



How to Think Music with Data

Translating from Audio Content Analysis to Music Analysis

Andersen, Jesper Steen

Publication date:
2017

Document version
Peer reviewed version

Citation for published version (APA):
Andersen, J. S. (2017). *How to Think Music with Data: Translating from Audio Content Analysis to Music Analysis*. Det Humanistiske Fakultet, Københavns Universitet.

HOW TO THINK MUSIC WITH DATA

Translating from Audio Content Analysis to
Music Analysis

JESPER STEEN ANDERSEN

Title: How to Think Music with Data

Subtitle: Translating from Audio Content Analysis to Music Analysis

Thesis submitted to the PhD school at University of Copenhagen, Royal School of Library and Information Science, June 23th 2017

Author: Jesper Steen Andersen

Word count: 67.670 excl. bibliography and appendices

Supervisors:

Jack Andersen, University of Copenhagen, Denmark

Morten Michelsen, University of Copenhagen, Denmark

Evaluation committee:

Jeppe Nicolajsen, University of Copenhagen, Denmark

Iben Have, AU, Aarhus University, Denmark

Alan Marsden, Lancaster University, UK

Contents

Links to Data.....	5
1. Introduction.....	7
1.1 Situating the Project Part 1: Big Data in the Humanities.....	7
1.2 Situating the Project Part 2: The CoSound Project.....	12
1.3 My Intentions.....	12
1.4 Premises.....	14
1.5 On Writing Style.....	17
1.6 Content of thesis.....	17
1.7 Terminology.....	19
1.8 Rounding off Chapter 1.....	22
2. Related Research.....	23
2.1. Research on MIR and data analysis.....	24
2.2. Related Musicological Research.....	27
2.3 Related Digital Humanities Research.....	36
2.4 Rounding off Chapter 2.....	40
3. Focus Points Where to Focus and How to Focus.....	41
3.1 MIR and Musicology - Different End Goals.....	41
3.2 Concern #1: How to Use Quantitative Methods in a Qualitative Discipline?.....	43
3.3 Concern #2: How to Interpret MIR Features?.....	47
3.4 How to Solve These Concerns.....	50
3.5 Choice of Analyses.....	52
3.6 Rounding off Chapter 3.....	53
4. From Audio Content Analysis to Music Analysis.....	55
4.1. Data Creation with ACA.....	55
4.2 The Advantages.....	59
4.3 How Humanities Objectives Fit with Data Approaches.....	65
4.4 But What to Learn about Music with ACA Methods?.....	69
4.5 Can We Trust Data?.....	78
4.6 Can We Trust the Analysis?.....	81
4.7 Can We Trust Data-Driven Approaches?.....	84
4.8 Rounding off Chapter 4.....	91
5. Echo Nest's Features Bridging from Machine Learning to Musicology.....	92
5.1 Introduction to Machine-learned Features.....	92
5.2. On Echo Nest and My Purpose.....	94
5.3 The Features.....	96
5.4 Introduction to the Values - the Basics.....	99
5.5 Epistemological Value – How to Interpret.....	103
5.6 Practical Value - The Usefulness of Reductions.....	110
5.7 Perspectives for Musicology.....	116
5.8 Rounding off Chapter 5.....	120

6. Then the Science Guys Entered the Room	121
6.1 The Analysis	121
6.2 My Purpose	126
6.3 Challenging the Epistemological Claims	126
6.4 My Interpretation of the Study	132
6.5 Prospects	134
6.6 Conclusion.....	138
7. A Corpus Study of 89 DJ Sets	140
7.1 Introduction	141
7.2 Step 1 - Exploring the Features.....	145
7.3 Step 2 - Surface Views: Exploring the Datasets by Mapping Them.....	166
7.4 Step 3 - The Shape of the Set: Analysis on the Macro level	176
7.5 Step 5 - Exploring Compositional Traits: Analysis on the Meso-level.....	181
7.6 Conclusion.....	188
8. Conclusion	193
8.1 What to win?	193
8.2 What to learn?	195
8.3 How to incorporate?.....	197
Appendices	201
Appendix 1. Resources (in Spring 2016).....	201
Appendix 2. Correlation Matrix for Echo Nest Features.....	204
Appendix 3. Loadings for Figure 51	205
Appendix 4. PC 3-4 for DJs and the Reference Corpus	206
Appendix 5. The journey of the average set.....	207
Table of Figures	208
Bibliography	210
Acknowledgements	222
Summary.....	224
Resumé.....	225

Links to Data

For the evaluation committee, I have enclosed a USB stick with all the relevant data, except Spotify playlists.

Chapter 5

Spotify Playlist:

<https://open.spotify.com/user/hviderusland/playlist/0uwDlprNHlMVZ816olkwTE>

CSV files containing the features used for analysis:

<https://zenodo.org/record/803950>

Chapter 6

Three songs mentioned during the chapter:

Spotify link:

<https://open.spotify.com/user/hviderusland/playlist/7mpXI3WTv6fHEIQ1wKOBV>

6.3.2

Spotify link:

<https://open.spotify.com/user/hviderusland/playlist/6sIMd9hEEYNZKiZqjUxqW2>

6.3.3

Spotify link:

<https://open.spotify.com/user/hviderusland/playlist/37BPzD6ISA4y44Y8VzRN7I>

6.5.2

Spotify link:

<https://open.spotify.com/user/hviderusland/playlist/1v4tqZvBcTCOdy7Ay4HLKM>

Mauch et al.'s dataset:

https://figshare.com/articles/Main_Dataset_for_Evolution_of_Popular_Music_USA_1960_2010_/1309953

Chapter 7

Reference Corpus:

Spotify link:

<https://open.spotify.com/user/hviderusland/playlist/0CuXAzHqCHuLtKUFZNHUtr>

DJ Sets:

At the moment of writing most DJ sets can be found at <http://www.mixing.dj> or www.1001tracklists.com. However, I cannot guarantee that the exact timing of these recordings matches the audio files used for the analysis.

Full dataset.

Containing Raw features, MATLAB scripts, Spreadsheet, and Tableau visualization files:

<https://zenodo.org/record/806207>

New and additional datasets

New and additional datasets created after I submitted the thesis, will be posted on:

<https://mirinmusicology.wordpress.com/2017/06/23/how-to-think-music-with-data-links-to-data/>

CHAPTER I

Introduction

1.1 Situating the Project Part 1: Big Data in the Humanities

In 2010, Google released the Ngram Viewer.¹ The viewer can plot how many time words or a string of words has occurred within 5 million books through the years, from 1500-2008. It enables us to instantly find out that “United States is” became more common than “United States are” from 1880, a few years after the civil war and onwards (Aiden and Michel 2013, 4). This tendency can be viewed as an indicator of American national identity. Another example is that the occurrences of “I” are increasing in recent years, compared to “we”. This trend has been viewed as a growing focus on the individual.² The occurrences of “women” now are on a level with “men”, whereas the word “men” occurred eight times as often as “women” in the digitized books dated from the 1800’s. And equally important, the Ngram Viewer allows anyone to type in own queries. The graph will pop up instantaneously.

Another example of the power of digital techniques is literary scholar Matthew Jockers who, a few years later, applied a computer for counting the occurrences of the most common words from a corpus of 106 19th century novels (Jockers 2013, 63-104). He divided each book into ten parts and took a sample of these 1.060 parts. Hereafter he trained the computer to trace correlations between the most commonly used words in the book part and the literary genre of texts. Next, he asked the computer to apply this knowledge to “guess” the genre of the remainder of the parts, observing that the machine was able to guess the correct genre for 67% of the parts. Jockers did similar queries with author, gender, decade of publication, discovering that the computer was able to guess the author correctly in 93% of the cases, gender 80%, and decade 53%. These results were applied to attempt to quantitatively grasp often-debated questions about how much different factors influences artistic outputs. How much is the time of writing apparent in a

¹ <https://books.google.com/ngrams>

² <https://www.information.dk/kultur/2014/01/vores-stedord-sladrer-hvordan-taenker>, retrieved May 23, 2017

text? How much is the author apparent? How much is the author's gender traceable in the writing style?

These two cases are a part of a general, growing trend within the humanities: Large corpora of cultural objects analyzed by means of digital techniques to assist humanities scholars answering questions about various aspects of cultural production that were practically impossible to investigate only 10 years ago. The first prerequisite for these types of analysis of culture is the large amounts of cultural objects that have digitized. Abundances of cultural objects in all genres are being digitized for archival projects and commercial purposes, and users from all over the world are digitizing and creating culture, sharing it on digital platforms. The second prerequisite is the development of digital methods that can analyze these cultural objects that have been digitized. A computer can detect and count which words that occur in a digital text. It is fairly good at detecting objects and colors in a digitized picture. And it can often guess the genre of a piece of music, if provided with an audio file.

Referring to digitization's effect on literary studies, Catherine Hayles has argued that "perhaps the single most important issue in effecting transformation is scale" (2012, 45). The growing amount of digitized cultural objects allows new queries that would have been practically impossible in the manual realm, as there are natural limits for how many cultural objects a person can engage with during life. There are upper boundaries for how many books one can read, or how many pieces of music one can listen to. However, digitized cultural objects can be searched and analyzed using digital techniques in various ways. And thereby they can allow scholars to grasp more objects. Therefore, the combination of digitization and digital methods has a potentially large effect on how humanities' scholars can conduct research. We are now "able to approach culture in a radically new way", as Berry has stated (2012, 2). Humanities' scholars now can investigate culture in ways that were practical impossible 20 years ago.

This situation within the humanities can also be seen as a part of an even broader technological and societal trend these years. The idea for this project was born in 2013, the year Gartner named *Big Data* (from now on in small cases) as one of the most hyped technologies.³

³ <http://www.gartner.com/newsroom/id/2575515>, retrieved January 5, 2017

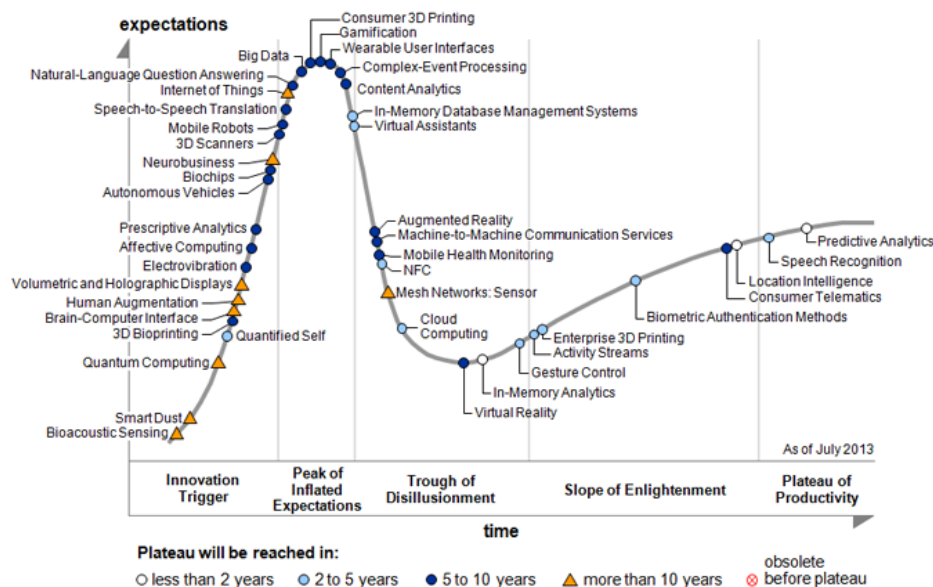


Figure 1 Gartner's Hype Cycle of 2013. Big Data reached the "Peak of Inflated Expectations."

Big data has been referred to as “things one can do at a large that cannot be done at a smaller one, to extract new insights or create new forms of value” by Mayer-Scönberger and Cukier (2013, 6)⁴. They thereby insinuate that large datasets can produce new knowledge and that there are many good reasons why we should exploit the potentials of large datasets.

The methods behind big data are primarily relying on knowledge derived from computer and data scientific fields, but they are applicable in many scholarly fields and gradually entering humanities’ research as exemplified above. Other humanities’ examples include Franco Moretti’s visualization of the characters’ relationship in *Hamlet* (2011). Or Lev Manovich’ social media project, *Inequaligram*, which traces in which tracts of New York locals and visitors respectively post pictures on Instagram (Indaco and Manovich 2016). Just to name a few.

However, it is no coincidence that the scholars and studies I have mentioned so far mostly have been based within literature studies despite this project concerns music. There are more good examples of large-scale text studies than large-scale music studies. I amongst

⁴ Though they also state that “[t]here is no rigorous definition of big data”. See also (Dutcher 2014) for a broad range of meanings of the term *big data*.

other ascribe this to a combination of more digitized texts, in conjunction with text being easier to process analytical with digital techniques⁵

The good news for musicology is that the circumstances are changing and becoming more equivalent to literature studies': Audiovisual material is now beginning to pile up more storage capacities. The music intelligence company Echo Nest's database of digitized music currently contains, for example, more than 37 million known songs.⁶ Plenty of online services contains huge amounts of digitized music, such as Spotify, last.fm or Soundcloud which hold both commercially released music and music distributed online by amateurs. In addition, various archival projects have digitized large collections of audio cultural heritage. One example is the Danish LARM project⁷ that digitized more than 1 million hours of Danish Radio and which indirectly led to this PhD project. Other digitization projects include UBUweb⁸ and Europeana⁹. Each of these corpora forms a potential new basis for posing a myriad of questions about music.

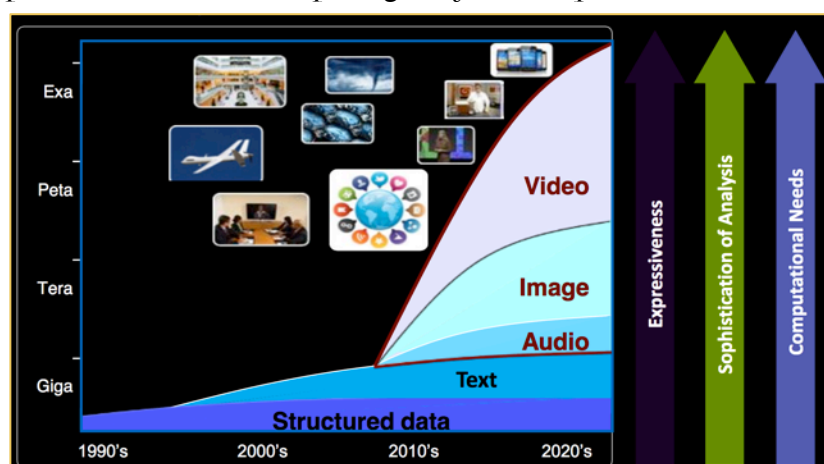


Figure 2 A growing share of digitized data is audio-visual data. This is a relatively recent trend. (Hessen 2014, slide 10, Quoting: IBM Market Insights based on composite sources / GTO 2013)

Parallel with the growing amount of digital audio, the field of Music Information Retrieval (MIR) has managed to develop tools able to calculate aspects of music from measuring on

5 As an indication of this, I did a search among the analyzing tools on Dirt, Digital Research Tools, a website that "aggregates information about digital research tools for scholarly use" (<http://dirtdirectory.org/>, retrieved November 2, 2015). 29 of the tools available on November 2, 2015 provided the possibility of working with text, while only 3 could deal with audio. 21 of these tools directly in their short description described that they could perform automatic text analysis task to some extent. None of the tools could, according to their short description, perform automatic audio analysis tasks.

6 This is a lot of music, even if you take into account that many of the songs are the same song in different audio formats. (the.echonest.com, retrieved January 9, 2017)

7 <https://larm.sites.ku.dk/>, retrieved May 18, 2017

8 <http://www.ubuweb.com/>, retrieved January 9, 2017

9 "Multi-lingual online collection of digitized items from European museums, libraries, archives and multi-media collections, with procedures for content providers." (<http://www.europeana.eu>, retrieved January 9, 2017)

audio files. I refer to these tools as Audio Content Analysis (ACA) tools (Lerch 2012). They open the way for releasing a similar potential as the other humanities' fields for deploying digital methods on digital objects. With all this information available and all this information that in addition now can be created automatically, musicology could stand to face a similar development as other humanities fields: Musicology could develop from being a data-poor into a data-rich discipline, as musicologist David Huron envisioned in 1999.

At the time of writing, there are small signs that this potential is beginning to become released. We have seen a few studies that exploit both the massive digitization of music and the power of ACA tools. Examples of these include (Mauch et al. 2015; Serrà et al. 2012; Echo Nest 2013a) which all have investigated large-scale trends in popular music's development. These studies provide overviews of music that would have been practically impossible in the manual world. And they have consequently also gained attention in the media. However, their origin and approach indicate that there is a gap between disciplines. The literature studies were all conducted by humanities trained scholars while the music studies, in comparison, primarily have been carried out by computer or data scientific trained scholars. One of the consequences is that the music analytical value of these studies is opaque.

More generally, there seem to be a gap between musicology and MIR, as several musicological scholars have advocated (Anja Volk and Honingh 2012, 73 provide examples of these): In 2005, Nicholas Cook declared that we were standing in front of a "moment of opportunity" for a closer collaboration between the two fields. But he also explained that the full potentials were far from being fulfilled. Marsden (2009) characterized that there was a gulf between computational and traditional approaches to music analysis. And Kranenburg et al. have diagnosed that one problem is that "[a]lthough [Musicology and MIR] both deal with music, there seems to be a gap in the ways of understanding it" (2007). When scrutinizing the large-scale music studies mentioned above, it illustrate well that there are many differences and discrepancies between humanities' and natural scientific approaches. However, the good news is that this does not automatically eliminate the opportunity of applying MIR as an instrument for approaching music in ways that are beneficial for musicologists.

1.2 Situating the Project Part 2: The CoSound Project

Apart from the trend of deploying digital methods for humanities' analysis, the other defining circumstance for this project was the larger research project in which it takes part, the Danish sound research project, CoSound¹⁰. One of CoSound's guiding hypotheses is in line with the prospects I delineated above. Namely that "[d]igital audio processing has yet to realize its full potential to enrich human communication, entertainment, and our cultural heritage."^{11,12} . Thus the CoSound Project can also be seen as part of this trend of applying digital techniques for humanities' purposes.

CoSound's vision was to "develop a flexible modular audio data processing platform for new products and services in the commercial sector; the public service sector; and in educational and cultural research."¹³ My project was associated with the last target group, the educational and cultural research. Consequently, my deliverables within the project were defined from the start. I had "to study and evaluate the automatic extraction techniques and results produced by CoSound, and to evaluate them in relation to real user needs and requirements."¹⁴ But since CoSound's results have been produced simultaneously as this Ph.D. project has unfolded, there were no other real users than me.

1.3 My Intentions

Due to the circumstances and trends delineated in section 1.1, I wanted to investigate whether there was renewed incentives for improving the link between MIR and musicology: I saw an unfulfilled potential of deploying digital audio processing methods in musicological research. Other humanities' fields had begun applying tools for analysis. Musicology was now facing similar circumstances; there now are vast amounts of digitized music and digital methods that can investigate it. Time was ripe for testing whether there was further basis for contributing to a similar development for musicology. I found that there was surplus of methodological literature on MIR methods, a lot of this very technical and task-oriented, conducted by engineers *for* engineers. But compared to e.g. literature

¹⁰ www.cosound.dk, retrieved May 18, 2017

¹¹ Digital audio processing is another word for audio content analysis

¹² CoSound Project Description - Appendix B, p. 2

¹³ CoSound Project Description - Appendix B, p. 2

¹⁴ CoSound Project Description - Appendix B, p. 6

studies there was a general lack of large-scale music analyses conducted by musicologists applying MIR methods to answer musicological questions.

1.3.1 - Five Dogmas

The setting and the predefined premises for the project lead me to set up some dogmas for my project. I wanted to investigate how to:

- A) apply MIR methods on audio files
 - B) on corpora larger than musicological researchers typically would investigate
 - C) consisting of Western popular music
 - D) for musicological purposes
 - E) And I wanted to conduct a case study
- A) The project should be limited to study automatic analysis of audio files and not symbolic music. Firstly, because audio files represent fuller sound data than scores, and thereby enables the analysis of more musical levels than scores. Secondly, it is easier for most people to get access to audio files than scores. Thirdly CoSound's primary object is audio.
- B) I saw that the one of the largest transformation that digital methods entail is that it now becomes possible to analyze large corpora of audio files.
- C) MIR tools are primarily developed and tested on western popular music. Therefore, I supposed that it would increase my chance of success to apply them on a corpus consisting of this type of music.
- D) Musicological purposes relate to the task predefined in the CoSound project description and match with my own competencies.
- E) There are more theories about how to connect between ACA and musicology than there are good examples.

1.3.2 Research Question

I extracted the research question that guides this thesis out of these dogmas. I wanted to investigate:

How can ACA methods be used for conducting large-scale analyses of audio files (of western popular music) for musicological purposes?

The *how* is implicitly indicating that MIR methods, in fact, can be used. If I had omitted the *how*, I would focus the investigation on whether or not MIR methods can be used. Many scholars from musicological disciplines have in recent years argued that they can (Honingh et al. 2014; Cook 2005; Kranenburg et al. 2007; Volk, Wiering, and Kranenburg van 2011). In addition, MIR has repeatedly demonstrated that audio signal processing methods can calculate or estimate many aspects of music.

Moreover, I have chosen the *how* to stress that there are many modes of applying the techniques. My focus will be on disseminating the prospects and pitfalls when using MIR methods. Thereby I am pointing the attention towards the following sub-questions:

- Which new questions can musicologists pose and answer music with ACA methods?
- What can musicologists learn from ACA assisted large-scale analyses?
- How can musicologists incorporate ACA methods into their practices?

1.4 Premises

1.4.1 Target Groups

The project is primarily an information scientific study. It concerns and studies issues that relate to data, the creation, management, and analysis of data. It studies these issues in relation to a humanities target group, musicologists, who are researchers who can benefit a lot from knowing about the methods. Thus, this study also has a pedagogical aspect, since it seeks to contribute with perspectives, prospects and critical awareness for musicologists, to guide the process of musicology developing into a data-rich field. I am aware that there are several sub-disciplines within musicology, and they cover separate areas, pose different questions, and apply different methods to answer them. My choice of dogma C), to focus on western popular music, to some extent restrict the scope of the study. However, at the

same time, I will seek to address general issues that relate to music analysis with digital methods. My own educational background is interdisciplinary,¹⁵ and one of its primary components is musicology. Therefore I have prioritized that my main center of attention is to bring my humanities' mindset into play when I investigate ACA and reflect on how to link from ACA to musicological knowledge.

Therefore this project will not be ground-breaking technical research. But this is not equal to claiming that MIR should not be interested in my findings. MIR is itself an interdisciplinary field (J. S. Downie 2003a), and many of its challenges could benefit from musicological knowledge (W. Bas de Haas and Wiering 2010). I have chosen music selection for my case study, as I hope as a possible side effect to inform MIR with analytical insight to one of its primary tasks, music recommendation (Celma 2010).

1.4.2 Elapsed time conducting audio content analysis is not considered

There are two main applications of this thesis: Either you want to understand an analysis created with ACA methods, or you want to deploy the methods for assisting your own analysis. In the latter case, the time it takes to conduct an audio content analysis will depend on factors out of my control. These factors include the state of the development of user-friendly interfaces or your coding skills, and they influence how much time will have to be spent conducting an analysis. Thus the practical concerns set the limit for how inclined a researcher will be for initiating an analysis. The time spent on something is time spent on not doing something else. And this something else could, for example, be to improve other competencies. Nevertheless, the general technological development gives me reason to believe that more user-friendly software than the current will be created. When this happens this thesis should be able to provide considerations, inspirations and guidelines on how ACA can be implemented.

1.4.3 External Premises

Coming from outside MIR into this very technical field that lacks user-friendly software, I was to a large extent on reliant on others' assistance. Therefore, the external collaborators in the CoSound project and the state of user-friendly MIR software played a crucial role in my choices of analysis subjects.

¹⁵ My master is combined by mathematics, statistics, musicology and cultural studies.

My initial idea was to practice large-scale analysis. I saw that the best way that I could investigate and demonstrate the power of tools was by getting hands-on experience with them. I agreed, and still do, with Nicholas Cook who had explained that we need “studies that are grounded in mainstream musicological problems and that make use of computational tools as simply one of the ways you do musicology” (2005, 1). I wanted to conduct a large-scale study of the music selection strategies in the Danish Radio. Such a study would exploit the large Danish radio archive, digitized by the LARM project, and this idea was formulated as my initial project proposal in my application. I (and the CoSound team) pursued this idea long way through my project time. However, due to both technical and juridical constraints too large to handle myself within the project time, this study stayed at the level of design but was never carried out. I had to change direction during the course.

While waiting to get my data from the radio archive, I was working on and investigating theoretical and methodological concerns. I was also engaged with some cases, which were chosen from a combination of what was possible and what I found enabled me to investigate my research question the most. At the beginning of my project, I was handed algorithms that allowed me to engage with how MIR features are calculated, and how they translate into musicology. This initial became a workshop paper (Andersen 2014), which I rewrote for Chapter 5. In the middle of my project, I was engaged¹⁶ with a big data music study questioning the epistemological value of it. This study activated my curiosity about my target group, musicologists, why did they dismiss the analysis? What could I learn about them from their criticism? I wrote about it for my Chapter 6. Towards the end of the project, I was provided algorithms that allowed me to upload audio files and retrieve audio features from it. This enabled me to pursue and complete parts of my original plan; to practice large-scale analysis with ACA tools. The findings are reported in Chapter 7.

The result is that this dissertation has been very much guided from what was available to me. This approach resembles a data analytic approach, where the learnings to a large extent are sought from the information that is obtainable. On the organizational level, the learnings are amongst other that there are practical restraints when being part of an interdisciplinary research project where each participant have their own incentives to pursue while contributing to the larger project.

¹⁶ Amongst other in Danish media: <http://www.radio24syv.dk/programmer/ak-24syv/11567882/ak-24syv-12-05-2015-1/> and <http://www.dr.dk/radio/ondemand/pl/pl-morgen-2015-05-09#>!

1.5 On Writing Style

It is *I* who conducts the analyses, and *I* am the one who writes this thesis. Therefore, I have chosen to write it in the subjective *I*-form. This choice is also a deliberate strategy to stress and enhance the subjective aspects of the analyses. They do not seek to be objective, although the results are objective in the sense of being reproducible, and although the methods are derived from disciplines with objective ideals. But I interpret the data. With this, I seek to emphasize that there are always subjective elements in the interpretation of data and that data analyses rarely are as objective as they may seem to some.

1.6 Content of thesis

The thesis consists of 9 chapters.

I will in Chapter 2 delineate what prior knowledge research I base this thesis upon. The chapter serves to demonstrate that thesis draws upon knowledge from different fields that traditionally are regarded far from each other. Firstly, I will delineate the achievements of MIR, which has enabled that I could realize this project. Next, I will point out musicological research that discusses how to apply MIR in musicology; the potentials, practical applications, and challenges that arise in continuation. Hereafter, I will provide examples of prior musicological empirical large-scale, but I will also manifest that there are many signs that the relationship between MIR and musicology is far from accomplished. Finally, I will delineate the digital humanities' theories I apply to assist me combining the fields I seek to combine.

Chapter 3 will continue along the path set up in Chapter 2. Its primary purpose is to diagnose where I see the most urgent points to investigate and to explain how I have chosen to do this. I will elaborate on the differences between MIR and musicology, in relation to purpose and culture. However, I will also explain why these discrepancies do not necessarily prevent that the one can draw benefit from the other. One of the most obvious current problems is that musicologists have no experience in understanding and interpreting ACA metrics, and I find this point crucial to discuss and improve in order to progress creating a better integration.

Chapter 4 constitutes the theoretical part of this thesis. In the first section of the chapter, I will ask what kind of data can be created? I will therefore shortly introduce to ACA and

what it enables us to measure in music: We can measure aspects very directly from the audio signal, we can simulate well-known musicological measures and it is possible to simulate human comprehensions of music to some extent. The emphasis is, however, put on the chapter's second part, which asks how musicologists can incorporate data into their practices. Firstly, I ask how more data in combination with digital methods can improve scholarship. Secondly, I ask how *musicologists* can apply digital methods for large-scale analyses. I will address issues that concern the types of questions humanities scholars pose, the lack of both music analytical and ACA standards and how to formalize music analytic inquiries. In continuation hereof, the delicate question of how much we can trust the data, and consequently also our analyses, arise. I will, therefore, line up what I see as the most crucial points that require special attention in relation to these questions.

Chapter 5-7 will consist of case studies that deal with important aspects of my topic. Chapter 5 will be devoted to the metrics provided by the music intelligence company Echo Nest's metrics. Though Echo Nest's primary purpose is to create algorithms for music recommendation, some of their means are interesting for my purpose. They apply amongst other apply machine-learned metrics to calculate intuitively comprehensible estimations of how music will be perceived. Therefore, I will investigate these metrics, how they relate to the music, and discuss whether there is a potential for applying and creating such metrics for music analytical purposes.

Chapter 6 will take as its point of departure in an already existing analysis, Mauch et al.'s "The evolution of popular music: USA 1960–2010" from 2015. This study concerns the creation of large amounts of data and the application of advanced data analysis techniques to handle this data. The researchers who conducted the study had a high degree of data scientific expertise and consequently the study exemplifies very well what data analysis enables us to do. How we can handle the data, make them manageable, visualize them, etc. However, at the same time, the study also exemplifies the epistemological limitations that arise as a consequence of the advanced data analytical techniques. The translation from data results to music analytical value impedes. The question that remains is what to do with these types of analyses, and how to progress from them?

As explained above, there is a general lack of good examples of musicological large-scale studies conducted with digital techniques. Chapter 7 will be an investigation of how much ACA methods can assist me analyzing a corpus larger than manually manageable. For that purpose, I will conduct a case study, an analysis of 89 DJ sets performed at the electronic dance music festival, Ultra Music Festival in Miami, 2015. I will apply ACA methods for the analysis of these DJ sets, with the goal of answering questions of musicological and

methodological relevance. The chapter will be worked out as a guided tour through all the steps in the analysis, demonstrating methodological concerns that arise along the way. The overall purpose is to arrive at conclusions about the music's acoustical qualities. The central question for this chapter is whether I can apply ACA methods to investigate what music the DJs' choose and in what order? Are there "formulas" that the DJ uses when programming the music, for example by contrasting low energy and high, or maintaining a monotonous expression? Are there any general tendencies in how DJs structure the courses of whole sets? Do DJs who play the same type of music apply similar compositional strategies?

The concluding Chapter 8 has the purpose of providing a broad summary of my study.

1.7 Terminology

This project mingles among different disciplines, and some of the concepts I apply are comprehended differently depending on discipline. I have chosen to settle definitions on some of the most crucial concepts to avoid misconceptions. These definitions work throughout the text. Below are listed some terms words that need further introduction.

1.7.1 Corpus - Metadata - dataset

To be able to follow the process of conducting the data analysis, it is useful to be able to discern between the *corpus*, *metadata*, and *dataset*.

Corpus (in plural corpora) denotes all the objects chosen for analysis. Corpus will most often denote the collection of audio files. Corpus is primarily derived from linguistics. Referring to Oxford Dictionary it means "[a] body or complete collection of writings or the like; the whole body of literature on any subject"¹⁷ or "[t]he body of written or spoken material upon which a linguistic analysis is based."¹⁸ However, the word is also used within musicology (Mauch et al. 2015; Anja Volk and de Haas 2013). If I strip the linguistic constraints from the latter definition of the two it says "the body of material upon which an analysis is based". This definition applies well to my application of the word.

17 3a (<http://www.oed.com/view/Entry/41873?redirectedFrom=corpus>, retrieved May 23, 2017)

18 3b, (<http://www.oed.com/view/Entry/41873?redirectedFrom=corpus>, retrieved May 23, 2017)

Metadata is “data that describes and gives information about other data”¹⁹. When analyzing audio files containing music this *other data* will often be equal to the corpus. However, to avoid confusions it might be worth noting here that there are various definitions of metadata, even within MIR. Li et al. define “Music Metadata” as useful data such as “artist name, track title, music description and data format” (2012, 5), a category that does not include the ACA methods’ calculations, the so-called *features* (see also 1.7.3). Lerch, on the other hand, includes features in his concept of metadata, which plainly denotes “data about data” (2012, 1). In this thesis, Lerch’s concept of metadata will be applied. Metadata will denote both all automatically extracted data and the humanly constructed metadata.

Dataset. I use this word to denote the full set of data applied for the data analysis. In practice the dataset is tantamount to the spreadsheets, I analyze: It is the dataset that I approach statistically to draw conclusions about my corpus. According to Oxford English Dictionary dataset means “[a] collection of data”²⁰, so the word could, according to this, and most other ways of using it, also be applied to what I have defined as a corpus. But in this case it does not; I simply found it useful to be able to discern and enforce a consistent terminology regarding these two.

1.7.2 Music Information Retrieval (MIR) - Audio Content Analysis (ACA)

Alexander Lerch (2012) discerns between MIR and ACA, I apply the same distinction:

Audio Content Analysis (ACA) is “the extraction of information from audio signals such as music recordings stored on digital media” (2012, 1).

Music Information Retrieval both denotes a task (to retrieve information from music) and a whole scholarly field that is centered around this task. Thus ACA is one of the techniques that MIR applies, while MIR also applies other music information retrieval tasks, such as “lyrics, user ratings, performance instructions,” handling of scores, etc. (Lerch, 2012). In this thesis, I will apply ACA when I want to make explicit that it is information derived from audio. I will apply MIR for more general statements.

¹⁹ <http://www.oed.com/view/Entry/117150?redirectedFrom=metadata#eid37413841>, retrieved November 26, 2015

²⁰ Oxford English Dictionary <http://www.oed.com/view/Entry/261122?redirectedFrom=dataset#eid>, retrieved November 26, 2015

1.7.3 Features (low and high level)

Features is the class of metadata which is automatically extracted from the audio signal. Features “include any acoustic propert[y] of an audio sound that may be recorded or analyzed” (Li, Ogihara, and Tzanetakis 2012, 5).

Low-level / high-level features. MIR divides features into low-level and high-level features. Low-level denotes features that are more directly extracted from the raw audio signal, such as spectral flux, ACF or cepstrum. High-level features are often compounds of low-level features and thus more complex, but they are often created with more regards to human interpretation and perception (Lerch 2012, 4–5). They include tempo, pitch, key, structure, etc. There is no clear boundary between low-level and high-level features.

Feature extraction refers to the process of calculating the desired features from the audio files by means of the algorithms.

1.7.4 Metric - Measure - (Metre)

While *features* are very concretely connected to automatic *audio content analysis* the words *metric* and *measure* denote more broadly applicable terms that relate to the handling of data more generally. However, both *measure* and *metric* have a different meaning within the fields I combine. Within music terminology they both relate to rhythmical metre. However, since both words are central concepts in statistics and data analysis, I have chosen to apply them in their data analytic sense. I will plainly avoid the use of *measure*, *metric* or *metrics* in relation to rhythmical metre. Instead, I will state it explicitly when it is the musical connotations I intend to evoke.

Measure (and **measurement**) need no further introduction, but **metric** has multiple meanings. In this thesis it denotes a “system or standard of measurement; a criterion or set of criteria stated in quantifiable terms.”²¹ There is a connection to low-level and high-level features: A measure indicates a more concrete, more directly obtainable and more objective attribute, and low-level features are likely to be classified as such. While a metric is a compound of measures²², thereby sharing similarities with high-level features.

21 Oxford English Dictionary, 4. (<http://www.oed.com/view/Entry/117657#eid37119435>, retrieved November 26, 2015)

22 https://samate.nist.gov/index.php/Metrics_and_Measures.html, retrieved November 26, 2015

1.8 Rounding off Chapter 1

In this chapter, I outlined the background and the driving motivation behind this thesis: The recent trends of applying data analysis techniques for humanities' purposes and the prospects in doing so. I also, however, outlined that there is an unfulfilled potential within audio-based research. In the following chapters, I will elaborate on how I see the current connection between MIR and musicology, where I identify the sore points, and where this study should focus to contribute to releasing the potentials.

CHAPTER 2

Related Research

In the previous chapter, I explained the setting and what lead me to set up the dogmas for the project. This chapter serves to situate my project in relation to the knowledge within the fields I combine. Its purpose is to demonstrate that I work within an interdisciplinary field and that this thesis consequently draws on knowledge from as various areas as data science, engineering, musicology, and digital humanities.

As explained in chapter 1, I saw most prospects of applying digital tools in relation to scale, so I wanted to examine how to apply ACA methods for enhancing the number of musical objects, we can study at once. One aspect of large-scale music analysis is that it requires some amount of data analysis, as Tim Crawford states in (Wiering and Benetos 2013). If I adapt the sort of judgment that according to John Tukey (Tukey 1962, 9 *al-a5*) is likely to be involved in almost every instance of a data analysis, it implies that a musicological data analysis requires knowledge about:

- The subject you analyze (the music).
- How the methods created the data, and what it represents (ACA).
- The particular data analysis techniques applied.

Roughly speaking, the disciplines of musicology, MIR and data science, respectively, produce this knowledge. These are also the fields I intend to combine, and consequently it is in the intersection between these fields this piece of research find its legitimacy. The fact that I combine such diverse areas of research at the same time implies that this will be a

crossover, interdisciplinary²³ study combining fields that traditionally are regarded far from each other, in relation to the methods, questions, culture and what to consider evidence. I will elaborate on this in Chapter 3. This thesis pursues the generation of synergies by engaging the breadth rather than scrutinizing in depth. I found that the field of digital humanities currently has the most elaborate and up-to-date theories about crossing over disciplines; from applying digital techniques, developed in the sciences to deploy them in a humanities' context. Consequently, this thesis also becomes a digital humanities thesis, more specifically a digital musicological one.

Hence, to answer my research question I will combine theories and practices from fields as diverse as data science, big data, engineering, computer science, information science, digital humanities, digital musicology, and musicology. In the following pages, I will explain what prior research I base my findings on, and what role they play for me. While doing so, I will also point to the gaps that I find most urgent to cover to combine the fields I intend. I will elaborate more thoroughly on the gaps in Chapter 3. In this chapter, I will commence by delineating prior research within the fields that have enabled the potential. Next, I will introduce the thoughts of previous musicological scholars who also have worked in the intersection I work. I will provide examples of prior analyses of large amounts of music. Finally, I will present digital humanities' scholars who more recently have thought about how to practice large-scale humanities' studies by applying digital methods.

2.1. Research on MIR and data analysis

As delineated in Chapter 1, the necessary conditions are here for conducting large-scale music analytical studies: The amounts of digitized audio are growing rapidly (Smith 2013) and the field of Music Information Retrieval has developed digital analysis methods capable of retrieving information from automatically analyzing these files. The latter of

23 A note on the choice of word: I use the word *interdisciplinary* to describe this study. This application is in accordance with Lin's description (2012). To Lin interdisciplinary is "referred to as an approach that allows researchers to work jointly and to integrate information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to tackle one problem" (298). Many projects and theories often note that interdisciplinarity is a central element of the digital humanities. However, Lin also remarks that "[f]uzzy definitions of these words mean that these categories are 'ideal types only' and serve mainly for theoretical discussions" (298). It is outside the scope of this thesis to contribute to this theoretical discussion about labeling concepts. I will therefore, for now, agree with the Digital Humanities Manifesto 2.0 which proclaim that "[i]nterdisciplinarity/transdisciplinarity/multidisciplinarity are empty words () unless they imply changes in language, practice, method, and output" (Schnapp, Lunenfeld, and Presner 2009). I emphasize that change is the important factor to investigate and pursue: My goal is to analyze the prospects of *change*, and therefore it is not within the scope to fill these various -disciplinary words with more meaning than written in this footnote.

these two is in focus throughout this thesis, and that is where I will commence this research overview.

2.1.1 Music Information Retrieval (MIR)

Music information retrieval (MIR) is a scholarly field that has as one of its chief activities to develop and investigate these digital techniques. At this time of writing, Wikipedia describes well the broad purpose of MIR: “Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music.”²⁴

Information from music can be derived from many sorts of sources, such as scores, lyrics, metadata, blogs, etc. (Li, Ogihara, and Tzanetakis 2012). I focus sheerly on the subset of these sources, audio files. MIR refers to these processes as *digital signal processing*, *audio feature extraction* or *audio content analysis*. These techniques allow for both “[p]reviously established non-digital research approaches [to] be ‘packed up’ into software programs” and new “methodological approaches that are intrinsically tied to the computer”, as Rieder & Röhle have expressed it (2012, 69-70). I will throughout the thesis refer these two types of approaches as born-analogue and born-digital, respectively. There is a wealth of technical research articles that describe the creation of these methods, evaluates them, or improves existing methods. Most commonly their focus is to determine whether algorithms can help solve a given task. Can computers, for example, assist the identification of the mood of a song if provided with an audio file? Or how can we create a better algorithm that can improve the precision on a certain task? However, the music analytic value is often not important and therefore not considered or taken into account in the articles. For the musicologist, one problem is that there often is a lack of experience in applying a given ACA method for analysis, and therefore it can be useful to rely on these technical, and formula-loaded, texts to get an impression of how the algorithms “think.”

Apart from the multitude of technical articles on ACA methods, (Li, Ogihara, and Tzanetakis 2012; Meinard Müller 2015; Lerch 2012) provide broader overviews of the techniques. George Tzanetakis have also presented some of the techniques at his UVic Music Information Retrieval course on Youtube²⁵ (George Tzanetakis 2014). Amongst the books, Li et al. present the widest scope, as it covers a broad range of topics from the most common audio content analysis techniques to social tags, hit song science and symbolic music. Müller demonstrates a narrower, and consequently also more in-depth, focus on

24 https://en.wikipedia.org/wiki/Music_information_retrieval, retrieved December 1, 2016

25 <https://www.youtube.com/user/gtzanetakis/videos>, retrieved November 24, 2016

music information retrieval sheerly from audio signals. Although Müller states that “the main focus of th[e] book is on computational rather than musicological aspects” (236), the book is well the best suited for also getting a basic musicological understanding of the methods. It covers from the basic of the Fourier analysis²⁶ of audio signals to a wide range of music processing methods, and it combines the music analytic with a computational focus. Lerch (2012) holds a more technical focus also covering only the automatic feature extraction from audio files. The book is, however, mostly oriented towards the complicated math behind the methods, or as Bob Sturm writes in a review: “The main contribution of the book is its collection in one source of the many features available for signal analysis. Most of these features, however, are presented without any reference to music audio content” (2013). From a musicological point of view, I regard Müller et al. (2011) as a better translator from audio content analysis to well-known music analytical concepts. This article demonstrates well that psychoacoustics is an important link for progressing from signal processing techniques to measures of music analytical aspects; of pitch, harmony, rhythm and timbre. At the same time, it also explains the challenges that arise when mathematical operations seeks to model human perception.

2.1.2 Data analysis

ACA methods can create a vast amount of information and data about any piece of audio. Consequently, the field of MIR is responsible for one part of the data analysis, the creation of data. However, I also rely on general data analytical texts, which concern the value of this data and how to handle it. The concerns proposed in John Tukey’s *The Future of Data Analysis* from 1962 are still useful for this context. The text focuses on how to apply data analysis as a tool to retrieve insights about a subject and not as a goal in itself. I also use Mayer-Schönberger and Cukier’s *Big Data* (2013) provides many arguments for the benefits of creating a lot of data. And they provide a large number of (primarily commercial though) examples of applications that can serve as an inspiration for what is possible when we have large amounts of data. Conversely, (boyd and Crawford 2012; Dalton and Thatcher 2014) are examples of big data critical texts that I apply as a counterbalance. Their main argument is that we now can analyze a lot, but we still need to interpret our data – on many levels. In a similar critical spirit, Rieder and Röhle (2012) discuss digital methods more generally in humanities, especially questioning their epistemological value and status. These texts will throughout thesis be applied to address concerns that relate to data analysis.

²⁶ Fourier analysis is the basic mathematic operation behind many of the features.

2.2. Related Musicological Research

2.2.1 Theories on computers in musicology and on large-scale studies

The idea of applying computers for analyzing music is of course not brand new. To my best of knowledge, (Bronson 1949) is the first example. Mendel (1962) and Erickson (1968) provide more elaborate theoretical reflections on the topic. In the 80's, Leonard Meyer (1989) argued for an approach that bears many similarities to what I today would call a big data approach. Enlargening sample sizes would allow musicologists to examine musical styles better. This argument later also proposed by (Cook 2005; A. Volk, Wiering, and Kranenburg van 2011).

Another important text regarding large-scale musicological analysis is David Huron's speech from 1999 entitled "The New Empiricism: Systematic Musicology in a Postmodern Age." Huron's main argument is that there is a relationship between the amounts of data at hand within a field and its methods, and this ought to concern our approach to evidence (2). For the case of musicology, he hopes that data richness will entail changes within the discipline, because it will require us "to embrace higher standards of evidence, and to be more acutely aware of the moral and esthetic repercussions of our knowledge claims" (177). However, I regard this discussion to somehow come after this thesis, and consequently, it is not my top priority. As I will argue later in this chapter, and elaborate further in Chapter 3, I see a need for establishing a proper link between ACA and musicology to enable the data-richness in the first place. After this, we can discuss the role of "standards of evidence." Besides, some of the ways music researchers can deploy ACA are not evidence-demanding. For example, if ACA is applied in an exploratory manner to create an overview of large amounts of music.

Three Nicholas Cook texts, one of them co-written with Eric Clarke, fit my mission better (Clarke and Cook 2004; Cook 2005; Cook 2010). They are to a larger extent discussing what musicologists can do rather than what they ought to. Cook and Clarke write:

"[e]mpirical musicology to summarize, can be thought of as musicology that embodies a principled awareness of both the potential to engage with large bodies of relevant data, and the appropriate methods for achieving this; adopting this term does not deny the self-evidently empirical dimension of all musicology, but draws attention to the potential of a range of empirical approaches to music that is, as yet, not widely disseminated within the discipline" (2004, 5).

This way of approaching music does not necessarily imply the use of digital methods but nevertheless resembles mine. It comprises the relationship between many of the key aspects of my analysis: large bodies of data, appropriate methods, potential and yet, not widely disseminated. The book also contains chapters on computational analysis and data collection; however these are more related to close reading practices, investigating individual works of music. Cook further elaborated this mindset six years later (2010), where he argues why more musical data is desirable, but at the same time maintains that a recursive relationship between listening and data is apt.

Where Meyer's and Huron's theories primarily are rooted in the analysis of scores, Cook specifically in (2005), a speech held at the annual ISMIR²⁷ conference in 2005, argues on the basis of computational tools that can analyze audio. His introductory line: "We stand at a moment of opportunity" is indicative of a spirit of wanting to release the prospects that could arise in the intersection between MIR and musicology. The mantra throughout the speech is that "musicologists are used to working with highly reduced data" (2). Firstly, the reduction regards the representation of data: "Notation is only an approximation, [...] dimensions like timbre and texture [...] aren't directly represented in notation at all" (3). And secondly, the reduction regards the scale of analysis. Audio content analysis tools enable us to diminish the reduction, as they provide an opportunity to work with larger datasets and with fuller sound data. I will add here that access is a further issue since a lot of music is easier accessible as audio than as scores. In addition, ACA methods reduce the music in ways that scores do not.

These texts by Cook fit well into my research question. They discuss the methods and their usefulness, but their emphasis is on how they can help achieve musicological knowledge. Cook and Clarke acknowledge that "musicology is or could be, in many instances, a significantly "data richer" field than we generally give it credit for" (2004, 4). This assumption permeates a lot of Cook's thinking on large-scale analysis, and consequently, it fits with one aspect of my object of study. At the same, Cook and Clarke insist on maintaining music analytic value. They do so, for example, by addressing the problems of prior use of "apparently objective analysis" in which the problem is "what one was meant to do with them, what their value was" (2004, 6).

27 International Society for Music Information Retrieval, www.ismir.net. See e.g. (Lee, Jones, and Downie 2009) for an analysis of ISMIR proceedings 2000-2008.

2.2.2 On music analysis

The traditional conception of music analysis needs a little rewriting and clarification to be made viable for large-scale analysis. According to Oxford English Dictionary *analysis* is defined as “[a] detailed examination or study of something so as to determine its nature, structure, or essential features.”²⁸ According to several characterizations of music analysis,²⁹ this “study of something” has traditionally implied one piece of music which takes “as its starting point in the music itself, rather than external factors” (Bent and Pople) and which focuses on examining the work’s internal structures.

It is almost needless to say at this point that music analysis in this context does not concern only one piece of music. However, ACA also takes its starting-point in the music itself. (This does not apply for all MIR methods, which can also retrieve data from other sources, such as artist metadata, people’s tagging, comments, ratings, etc.). Thirdly, the structures that traditional music analysis asks about in one piece now have to become quantified in order to become components in the analysis of many pieces of music. They consequently have to be understood and expressed in new ways, in terms of statistics.

ACA can, for example, estimate the number of chord changes, and thereby measure one aspect of the structures in a piece of music; a measure that teaches us something about the piece’s internal coherence. Or for example, the average energy level of the output (RMS) will often be of limited music analytical value, but the standard deviation of RMS can teach us something about the internal dynamics of the music, on either micro-level or macro-level, depending on how you summarize the measurements. In both cases, aspects of the piece’s structure will be represented by some numbers. When scaling up, these methods can thereby teach us something about the inner relations of components within

28 <http://www.oed.com/view/Entry/7046?redirectedFrom=analysis#eid> retrieved January 14, 2017

29 Oxford Music Online refer to *analysis* as “that part of the study of music that takes as its starting-point the music itself, rather than external factors. More formally, analysis may be said to include the interpretation of structures in music, together with their resolution into relatively simpler constituent elements, and the investigation of the relevant functions of those elements. In such a process the musical ‘structure’ may stand for part of a work, a work in its entirety, a group or even a repertory of works, in a written or oral tradition.” (I. D. Bent and Pople, n.d.)

In their reference guide to musicologies’ key concepts, (Beard and Gloag 2016) open their reference on “analysis” by explaining that: “Analysis is a subdiscipline within musicology that is concerned with a search for internal coherence within a musical work. It therefore takes the musical text – usually a score [...] - as the primary, autonomous object of study, focusing on an examination of a work’s internal structure” (13).

Nicholas Cook’s generalizes what music analysis does: “There are a large number of analytical methods, and at first sight they seem very different; but most of them, in fact, ask the same sort of questions. They ask whether it is possible to chop up a piece of music into a series of more-or-less independent sections. They ask how components of the music relate to each other, and which relationships are more important than others. More specifically, they ask how far these components derive their effect from the context they are in.” (1987, 2)

many pieces. Does artist X's music change chords more often in the 2000's compared to the 2010's? Is the dynamic variety growing? Etc.

Bent and Drabkin wrote that music analysis "is the means of answering directly the question 'How does it work?'" (1987, 5). I see a similar end goal of why we should apply the tools. They can help us answer 'how do the many pieces of music work?' But at the current and not very advanced stage of applying ACA methods in musicology, I am equally interested in investigating 'How do the methods work?' The testing of the methods is one of the dogmas in this thesis (see Section 1.3). They set the limits, and form the boundary of the scope of my analysis. They will be tested to find out whether they can help us find out how the many pieces of music work.

But a piece of music can work in many ways, and on many levels, and this is also why there is no fixed recipe for conducting music analysis. Therefore I will apply and modify the data analysis in a heuristic, ad hoc manner: From what is possible and feasible, and from what makes the best sense when taking practical considerations, the music and the questions about the music and methods into account.

2.2.3 Digital musicological theory

Besides the texts that discuss theoretical concerns connected to large-scale analysis, there are other matters in the intersection between digital methods and musicology, which have been addressed. These issues can roughly be divided into those which concern computational analysis of music, and those that concern MIR viewed through musicological glasses. I will briefly mention examples of these below, but I will return to these texts throughout the thesis, because they all, more or less en passant, include considerations that are useful for me and my analyses.

Nettheim (1997) has written a bibliography of statistics in musicology, though it covers the analysis of scores. Huron (2013) discusses the role of applying statistics in the age of big musicological datasets; now it has become possible to analyze whole populations instead of having to rely on sampling. This text is largely consistent with his speech from 1999, mentioned above, and likewise, its discussions are slightly off my question. Huron's focus is on the moral scientific implications of applying big data, but some of the warnings presented are useful for establishing good practices when working with big datasets. For example, some caveats arise when you informally explore a dataset by means of data analysis techniques. For example, Huron addresses that we have to discern between post

hoc theories and hypothesis testing when we report our results. Similarly, Wallmark (2013) argues for musicological engagement with big data, but for different reasons than Huron. Wallmark explains amongst other that it will enable musicologists to better engage in the public agenda. Relating to music analysis Anagnostopoulou and Buteau (2010) argue that it is impossible to conduct a neutral music analysis, although we apply computers that seem neutral. Marsden (2016) discusses ontology and epistemology of computational analysis.

In the latter methodological category are W. Bas de Haas and Wiering (2010), Wiering (2009), and Wiering (2007) who all discuss challenges that relate to MIR. Aucouturier and Bigand (2012; 2013) focus on the relationship between MIR and music psychology, especially on the difficulties in establishing a fruitful relationship between the two fields. They identify a couple of aspects that impede a fruitful dialogue. Kranenburg et al. (2007) discuss how to integrate MIR and folk song research. And Volk, Wiering, and Kranenburg (2011) represent a wider view and discuss chances and challenges of applying computational methods for music analysis, historical musicology, ethnomusicology and cognitive musicology and musical performance research. Honingh et al. (2014) seek to strengthen interdisciplinarity by providing four concrete musicological questions that could be examined by using MIR.

2.2.4. Data in use - empirical large-scale examples

Apart from the last-mentioned reference, these digital musicological articles focus almost entirely on theoretical concerns. I have not found many empirical studies that fit my criteria of applying audio content analysis methods for large-scale analysis for musicological purposes. Most of the large-scale studies that hitherto have been conducted have been created on either the basis of scores or manual listening. Of these studies, only a few deploy ACA-methods, and as I will demonstrate especially in Chapter 6, their music analytic value is complex and opaque. Although we are currently at a stage where audio-based research now is a "serious possibility" (Wiering 2012), there is a lack of good examples of studies. In this section 2.2.4, I will, however, line up examples of previous large-scale analyses of music to provide an impression of which questions researchers have posed.

Score analysis

Arthur Mendel's Josquin Desprez studies (1969) are some of the earliest large-scale studies conducted with the help of the computer. More recently, Vos and Troost (1989)

have reported that within their corpus, which contained both western classical and western folk music, they found a tendency that intervals larger than a large third most often ascend, while smaller intervals descend. Huron (1996) examined 6512 folk songs by applying the Humdrum Toolkit analysis software (Huron 1995) finding that the average melodic phrase was arch shaped. Huron has assigned a chapter for statistical properties of music, mainly for music psychological purposes, which includes many examples of large-scale studies (2006, 73-90). Amongst more recent studies is (Rodriguez Zivic, Shifres, and Cecchi 2013), which deploys more advanced statistical techniques to analyze melodic intervals in western classical music from 1730-1930. It thereby exploits the large collection of digitized scores in the Peachnote dataset (Viro 2011). Interestingly, their mainly data-driven techniques automatically formed clusters that correspond with the prevailing categorization of classical music periods.

Manually annotated

Amongst large-scale studies using a combination of manual and computer-assisted annotation is (Schellenberg and von Scheve 2012) which measured tempo and mode in Billboard Chart's top 40. Their findings suggest that there has been an increasing amount of mixed emotional cues, indicated by more fast songs in minor mode. Another example includes De Clercq and Temperley (2011) who have conducted statistical studies on harmonic structures in songs from the Rolling Stone magazine's list of the 500 Greatest Songs of All Time. They annotated this corpus manually and after that applied computer for statistical analysis of harmonic structures. Two years later they extended this analysis further by adding melodic transcriptions to the same corpus and performing further more advanced statistical calculations on it (Temperley and de Clercq 2013). These publications mainly consist of descriptive statistics. Another examples of analyses that use manual annotations is (J. A. Burgoyne, Wild, and Fujinaga 2011) which in conjunction with the creation of the Billboard dataset counted occurrences of chords, finding that more than 50% were major triads. Choudhury, Bhagwan, and Bali (2013) have measured the evolution in the distribution of melodic scales in Bollywood music from 1945-2013.

These studies have basis in symbolic music, scores, which traditionally have been the analytic object in musicology. However, to my best of knowledge ACA is not yet³⁰ very well capable of separating sources and transcribing each voice separately from an analysis of an audio file (Plumbley et al. 2002). There are many ways to statistically and

³⁰ This is a problem that I know have MIR's focus. But I am not updated on the current status of the solving of this task. The most recent status I have had on this task was this presentation from 2013 in which status and challenges were outlined. http://www.cosound.dk/wp-content/uploads/2013/07/MSOSAM-DigAudioCPH-Jun13_mark-plumbley.pdf - retrieved February 23, 2016. During my project time, I was not presented software able to separate sources.

automatically investigate tonal aspects (Nettheim 1997), but as audio content analysis encompasses many other methods than tonal and have proven successful in measuring music from other perspectives, I also want to include and focus on these.

Large-scale ACA-analyses

Measured in corpus and dataset size (Serrà et al. 2012; Mauch et al. 2015) are to the best of my knowledge the largest scientific studies that apply digital methods for the sake of analyzing the development of Western popular music. While (Echo Nest 2013b) is another similar example though not scientifically reported. These three studies have all reached remarkable attention in the media (e.g. Dredge 2013; Campbell 2012; Akpan 2015). However, not surprisingly the music analytic value of these studies is more complicated than reported in the media. On the one hand, the studies provide facts about the development of western popular music. While on the other hand, it is complicated to account for what these facts actually reveal about the music. I will comment more thoroughly on Mauch et al.'s analysis in Chapter 6, where it will form a basis for a discussion on how these studies can benefit musicological research, what to interpret from them, and their limits.

At a smaller scale, but still large-scale compared to traditional musicological studies, Honingh et al. (2014) applied MIRtoolbox for analyzing 683 songs in the Billboard dataset (J. A. Burgoyne, Wild, and Fujinaga 2011). The method was linear regression calculated for some MIRtoolbox features as a function of year the song appeared on the chart. In contrast to (Serrà et al. 2012)³¹, they did not find a connection between amplitude and year. A similar data-driven approach is found in Balen et al.'s (2013) study of choruses in the Billboard dataset. What is especially interesting in this context is that both studies apply born-digital ACA features. These serve as inspiration for my case study. In their starting point (creating a lot of data about music), and their objects of examination (the features) these studies resembles mine. However, both studies mainly expose methodological curiosity of how the metrics "behave" in relation to the music and each other (see Figure 3). Both the newly developed metrics and their interdependencies are useful background knowledge for my case study. However, neither of the studies focus on musically explaining the results.

31 And (Echo Nest 2013b)

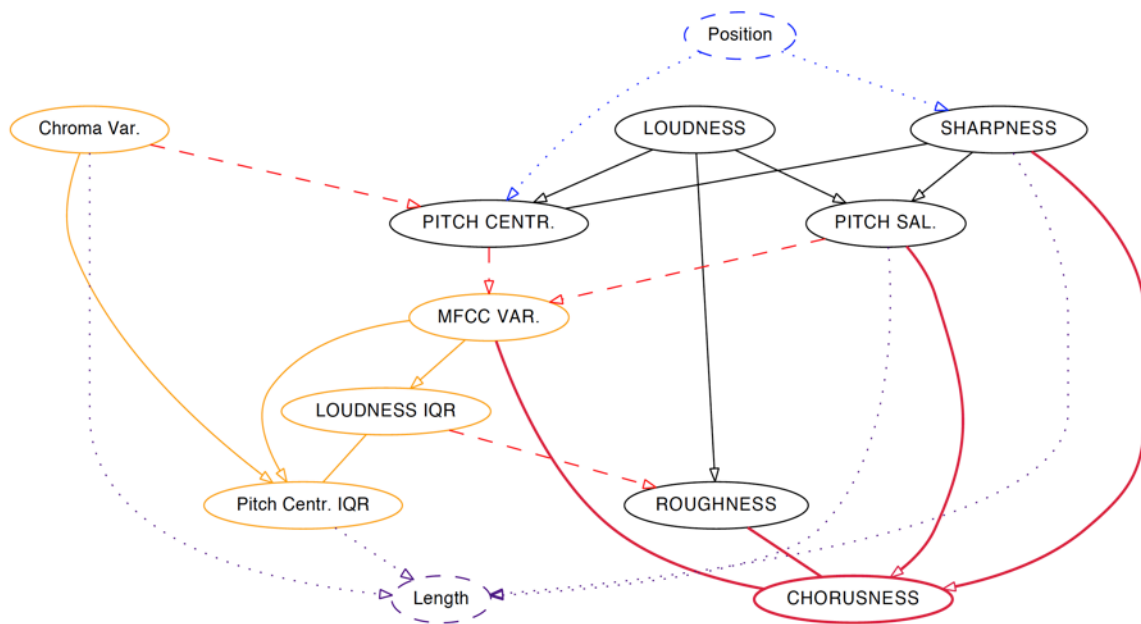


Figure 3 Balen et al.'s diagram of how the born-digital metrics depend on each other. Shown by a probabilistic graphical model (PGM). Arrows indicate dependency, e.g. Loudness influences Roughness as indicated by one black arrow. Colors do not have any mathematical value, but do have "thematic" value (Balen et al. 2013).

Data-driven approaches

In the data-driven end of the spectrum are examples of researchers who base their conclusions from analyzing machine learning techniques' efficiency. Esparza, Bello, and Humphrey (2015) investigated how well a computer, provided with data on rhythmical aspects, could predict genre from a corpus of Brazilian music. They hereafter used the results for considerations on how much each genre is characterized by its rhythmical aspects. Existing musicological theories were held up against the results showing coherence with them. Other examples for inspiration, though they do not apply ACA methods, are found in the realm of literature studies. As mentioned in the introduction, Jockers (2013, 63-106) applied a similar approach, measuring the accuracy of a model that automatically should predict author, book, gender, genre and decade. This approach was an attempt to quantitatively grasp often debated literary issues regarding strength of different signals: How much is the author's gender traceable in the writing style? How much is the decade of writing detectable? Fell and Sporleder (2014) have investigated to what extent musical genre can be recognized from the lyrics alone. It was possible: Rap was easiest identifiable, while folk music the least. Additionally, their model was surprisingly good at determining whether a song was high or low rated, and when it was written. In this context, as I will conduct a dance music analysis in Chapter 7, I will also mention that Herremans, Martens, and Sørensen (2014b) has demonstrated that Echo

Nest's features can be used for predicting what dance music becomes a hit. However, this study has a strong focus on model building and, as with much MIR research, it is difficult to retrieve music analytical insights from it. I do not intend to perform any similar studies, but I mention these as an inspiration of what digital methods also can do. In Section 4.7, I will comment on the music analytic value of data-driven studies.

2.2.5 The status of MIR in musicology

Despite the prospects of digital methods delineated in the previous pages, there are many indicators that these have not been succeeded with remarkable changes within musicology. Computational musicology is generally regarded as a detached branch of musicology (see A. Volk, Wiering, and Kranenburg van 2011 for references to other music theorists who agree). In 2005, Nicholas Cook enunciated that we had been standing at a "moment of opportunity [for a closer relation between MIR and musicology] for quite some time now" (1). He thereby hints that the full potential of a stronger collaboration was yet far from released. Two years later McKay & Fujinaga reported from ISMIR 2007 that there were "only a few musicologists and almost no music theorists have become involved with ISMIR to date". Alan Marsden in 2009 counted the number of musicological articles applying computational techniques. The enumeration lead Marsden to conclude that "there is little evidence that the gulf between 'traditional' analysts and those who use computer 'as their chief research tool' is narrowing" (137).³²

This feeling of an unreleased potential was also one of the main *raison d'être*; why this branch of the CoSound project was initiated in the first place. In its funding application from 2011, it was stated that "[d]igital audio processing has yet to realize its full potential to enrich human communication, entertainment, and our cultural heritage". And a similar recognition from the technician side "despite the fact that higher cognitive representations of audio are well-developed, they are not easily articulated or shared by non-experts." The CoSound project can also be seen as a part of a larger trend these years. Parallel with a greater focus on digital methods in the humanities generally, there have been similarly small signs of more musicological projects applying digital methods. In the report from the

³² Marsden's basis for this statement was a counting of the number of articles in musicological journals that either reflects on or directly applies computational methods. In *Music Analysis* 6 out of 221 articles did so. Four of them were mostly theoretic, and one was a manual analysis which makes a comparison with a prior computer analysis, while only one directly applies software for analysis. In *Journal of New Music Research*, 18 out of 365 fell into this category. Only a few of them analyses sound, there were no large-scale analyses, and they are all focusing on methodological concerns rather than music analytical. Many of these articles applied advanced statistical methods, such as Markov models or Bayesian statistics, implying that the majority of music researchers are not able to understand neither the intermediate calculations nor the conclusions (137).

Digital Musicology session at ISMIR 2013, digital musicology was considered a "growing field" measured in "a large number of [digital musicological] relevant papers" (Wiering and Benetos 2013, 1). This optimism or signs of improvement could be indicators of many tendencies. An improved relationship, mutual curiosity, or perhaps it is due to research political priorities? For now, though, it is not important which of these are most prominent. It is more urgent to address that there nevertheless still is a lot of ground to cover before the two fields are integrated. Before musicologists apply ACA methods seamless in their research.

2.3 Related Digital Humanities Research

This dawning of an improved relationship can also be seen in the light of a growing general acceptance of digital methods in the humanities. The field that preceded digital humanities³³, Humanities Computing has been relatively isolated from mainstream humanities (Finnemann 2015, 317). But digital methods are gaining acceptance and slowly becoming an integrated part of the humanities, and humanities computing is evolving into digital humanities.³⁴ As I depicted in Chapter 1, musicology is now slowly facing the same digital conditions as other digitized humanities' disciplines. Computational musicology could develop into digital musicology (Wiering 2012).

Consequently, I found it useful to widen my scope from sheer musicology into including other humanities' discipline, because the broader field of digital humanities could provide me with more elaborate reflections on how to practice humanities with digital methods, and with a larger catalog of studies for inspiration. One reason is that digital humanities theorize on the basis of very recent technological advances. These advances have implied that large-scale analysis using digital techniques has become a lot less time consuming and therefore more realistic to pursue. As this study focusses on large-scale analysis, I will narrow to digital humanities' theories that discuss aspects of digital methods and scale, or big data as a resource for humanities' research.

33 Digital humanities is a very loosely defined term, covering a very broad variety of humanities' actions. For a more comprehensive range of definitions see (Gold 2012, 69), or <http://dhdebates.gc.cuny.edu/debates/text/40>.

Or as Stephen Ramsay has put it: "Nowadays, the term can mean anything from media studies to electronic art, from data mining to edutech, from scholarly editing to anarchic blogging, while inviting code junkies, digital artists, standards wonks, transhumanists, game theorists, free culture advocates, archivists, librarians, and edupunks under its capacious canvas" (2013, 239)

34 While I regard it as out of the scope here to discuss eventual differences in the mindset between the humanities computing and digital humanities, I will rather note that digital methods in the humanities are gaining acceptance and are not as isolated from mainstream humanities anymore.

Ian Foster (2011) has from a technological viewpoint described what computation has brought along, mainly for the natural sciences. He argues that the humanities now are facing similar circumstances. Foster explains how computation can contribute to these fields by

- 1) creating increased access to objects
- 2) enhancing our perception (so we can find things that we would not find manually)
- 3) automating analysis
- 4) creating new opportunities for modeling and simulation.

3) is a pivotal point for my investigation, while 4) also will be investigated to the extent that modeling contributes to our understanding of music. I will for example in Chapter 5 examine Echo Nest's features which are based on models' calculations of properties of music. I regard 1) as a premise for this thesis. 2) is a consequence or effect of the others.

Manovich (2012) is more narrowly focusing on the benefits of the wealth of information about culture that is created. He provides a number of good reasons why humanities scholars should apply digital methods for analyzing them, and how they should do it. For this case, it is especially relevant that the text's starting point is an analysis of pictures. Pictures have similarities with music because digital methods can be used to describe properties that eschew words: With digital methods, we do not necessarily have to annotate the music to find stylistic characteristics. Further more the text is inspirational, amongst other because it directs the attention to the many types of data that constantly is created on the Internet. Youtube is one example that not only contains a wealth of music but also contains user comments, allowing insights into users' receptions.

Huron and Foster are not the only digital humanities' scholars who find inspiration in other disciplines outside the humanities. With the emergence of digital humanities a bunch of new concepts, many of them terminologically rooted in the quantitative sciences, have been created. To some digital humanities' scholars there is a spirit of standing on a threshold to a new paradigm³⁵ within humanities' research, and these new concepts are indicators hereof. Examples include *Cultural Analytics*³⁶ (Manovich 2015) encompassing “[t]he analysis of massive cultural data sets and flows using computational and visualization techniques”. The “goal is start systematically applying [“statistical data

35 Manovich (2009, 5) explicitly calls it a “paradigm”

36 The term *analytics* is here closest affiliated with Oxford English Dictionary's definition b): “The collation and analysis of data or statistics, esp. by computer, typically for financial or commercial purposes; the data that results from this; (also) software used for this purpose.”

(<http://www.oed.com.ep.fjernadgang.kb.dk/view/Entry/273413?redirectedFrom=analytics#eid>, retrieved February 1, 2016)

analysis, data mining, information visualization, scientific visualization, visual analytics, simulation and other computer-based techniques”] to the analysis of contemporary cultural data” (Manovich 2009, 6). *Culturomics*³⁷ (Michel et al. 2011) is another example. It refers to "the application of high-throughput data collection and analysis to the study of human culture". In this context, I apply these texts as inspiration.

For this thesis, I especially found inspiration in Matthew Jockers' (2013) thoughts formulated with yet another new concept, *macroanalysis*³⁸, , inspired by another data-rich field, economics.³⁹ Jockers' and my starting point of investigation resemble each other: We both investigate how to conduct large-scale analysis for humanities' purposes by applying digital methods. And Jockers, like me, approaches the application of computational methods for literary studies in a very inclusive manner.

Jockers provides examples of how to analyze a large, digitized corpus of properly metadated literature. He demonstrates statistical techniques and shows examples of questions that can be examined this way, for his case often-debated questions about for example style and authorship. Additionally, the book contains theoretic thoughts on how digital methods can enrich humanities scholarship and the role of large-scale analysis. Jockers elaborates on *why*, *how* and *what* to "read" with digital tools.

- *Why we should read with digital tools*

Jockers state that "[t]he goal of science, we hope, is to develop the best possible explanation for a phenomenon." Digital methods can for this purpose provide a means of gathering more evidence. They form a way to deal with the abundance of cultural objects now at hand and examine larger trends than possible before. They can, for example, help us overcome the limitation of often reading only canons (see also Moretti 2005 or Wilkens 2012).

My research question begins with a *how*, and therefore the question of *why* is not first priority to cover. The *why*-question comes both before and after this thesis. Before, because I found that the reasons why to conduct large-scale analysis were compelling.

37 The postfix *-omics* is a neologism that “informally refers to a field of study in biology ending in *-omics*”, wikipedia. (<https://en.wikipedia.org/wiki/Omics>, retrieved February 1, 2016) Or see <http://www.culturomics.org/Resources/faq> where Aiden and Michel in the FAQ, answer the question “Why did you call your approach 'culturomics?’” by referring to biology. (retrieved October 31, 2016)

38 Not to be confused with the music theoretic term macro analysis, a method of transcribing chords. (Wikipedia, https://en.wikipedia.org/wiki/Macro_analysis, retrieved August 10, 2016)

39 Macroeconomics studies entire economies, focusses “on the larger system” in contrast to microeconomics “which studies economic behaviour of individual consumers and individual businesses” (Jockers 2013, 24). In literature the most common way of reading is the close, hermeneutic reading of singular texts, which can be compared to microeconomics.

Thus the answer to the question forms this thesis' *raison d'être*; it was the reason to initiate my study in the first place. And after, because when I finish, I will be able to inform this *why*-question with practical music-specific concerns that can nuance the question. Notwithstanding, the question of *why*? is of course in the back of my head throughout my reasoning. The *how* is influenced by the *why*; the way we do things are guided by the reasons why we do them.

How to read with digital tools?

Jockers also elaborates on the role of interpretation and the collaboration between man and machine. One of the topics he covers is the relation between close, hermeneutic reading and the views that distant reading allow (see also Moretti 2013). Most digital humanities scholars emphasize that machines are not substitutes to human, but instead argue for a recursive relation between the machine and the human. This discussion is further elaborated by Hayles (2012).

For this thesis, I too found a need for approaching the machine humanly, for example, to be able to shuttle between close and distant reading. Not only because my target group is musicologists who wants to arrive at qualitative statements, but also because MIR calculate features and create statistics, but we have limited experience in interpreting the numbers. As I will argue in chapter 3, I found this lack of experience to be the most noticeable gap to cover in relation to examining my research question.

What to read with digital tools?

One side of this question is instrument-specific and relates to the efficiency of ACA. Jockers does, of course, not cover this as he studies literature, not music. The other more general aspect of the question regards the role of evidence in humanities studies. How humanities scholars pose questions and the types of answers that interest them. Jockers warns against "quantitative arrogance" and to present quantitative results as "definitive statements" (2013, 30). Rieder and Röhle (2012) discuss more elaborately on digital methods and their epistemological status and value. This question of what to relates to understanding my target group: What are their needs? And, equally importantly, what can they see with ACA tools?

2.4 Rounding off Chapter 2

In this chapter, I have outlined research that has inspired me, and that relate to aspects of my research object. MIR has created the possibility of creating music analytical useful data from audio files. Musicologists have thought about how to apply MIR technologies in their research and identified music-specific concerns in relation to this. And Digital humanities' scholars have provided me with more recent and elaborate theories and more practical examples of studies on how to apply digital methods in humanities' scholarship. These are consequently also the fields that I combine in this thesis.

Whitehead has stated that "[c]ivilization advances by extending the number of important operations which we can perform without thinking about them" (1911, quoted by Foster 2011). Computer scientific advances do not only allow musicologists to automate operations that previously would have been very time consuming, but they also enable operations that were impossible to pursue or hardly imaginable before computers. At least in theory. For it is currently difficult to determine to what degree these theoretical advantages transfer into practice, due to the lack of good examples of connecting audio content analysis to music analytical purposes.

There has been a cultural divide between humanities computing and mainstream humanities. As I will discuss more closely in the next chapter, there is a comparable cultural divide between MIR and musicology. Most notably for my case, this divide regards quantification and the notion of proof. However, Nicholas Cook, Matthew Jockers and other digital humanities' scholars have stirred up traditional ways of opposing quantification and interpretation. It is not necessarily the one or the other. Rather we should be open for that the former can inform the latter. We can now analyze large corpora of music with quantitative methods. But even if we can find statistical evidence of something in a corpus, the music analytic value of it is not necessarily self-evident.

In the next chapter, I will explain how I have chosen to focus my research. I will also explain what ideals and delimitations I have chosen in the pursuit of answering my research question.

CHAPTER 3

Focus Points

Where to Focus and How to Focus

In the previous chapter, I outlined the research I have found most relevant to employ in the pursuit of answering my research question. I explained that there is a gap between the fields I combine; most importantly there is a lack of collaboration across the disciplines, I seek to connect. The purpose of this chapter is to explain how I have chosen to tailor my research, and why it is so. Accordingly, I will in this chapter explain my mission and how I chose to delimit the topics covered in this dissertation.

3.1 MIR and Musicology - Different End Goals

The first and perhaps most obvious discrepancy, I will address, concerns the general level of purpose and goal: Although we have all these feature extraction techniques, we should keep in mind that “MIR is strongly task oriented” (J. A. Burgoyne, Fujinaga, and Downie 2015, 213). Therefore MIR’s primary goal is not to develop theories on music (W. Bas de Haas and Wiering 2010). Music analysis is only one of many MIR tasks, and in most cases, it is only a subsidiary goal to reach other more application-oriented goals, such as music recommendation, genre detection, emotion estimation, etc. Therefore there is an odd, non-homogenous relationship between the topics I investigate: Can musicologists utilize instruments that were not initially intended for them? The theoretic answer to this question is of course, ‘yes.’ History has seen many examples of applications of inventions that were not initially intended for the purpose. And scholars have accordingly argued that even though there often are commercial intentions behind MIR, it can still provide valuable insights for musicology (Anagnostopoulou and Buteau 2010; Marsden 2009).

Below is a coarse diagram of how MIR (the brown arrow), musicology⁴⁰ (blue), and I (the red), respectively, deal with different research components. In most cases, the elements

⁴⁰ The kind of musicology I am comparing most of my work with. I have excluded many branches of musicology, such as music psychology

will be more intermingled, but notice especially how the end goals differ: Even though MIR and musicology share some interests their end-purposes rarely match. This discrepancy affects both the development and the application of the individual elements in the diagram. I will implement the tools that MIR often use, but for a purpose similar as musicology's. The thin brown arrow ending in "Facts" is representing MIR assisted music studies most often conducted within MIR methods, I will discuss one of these in Chapter 6.

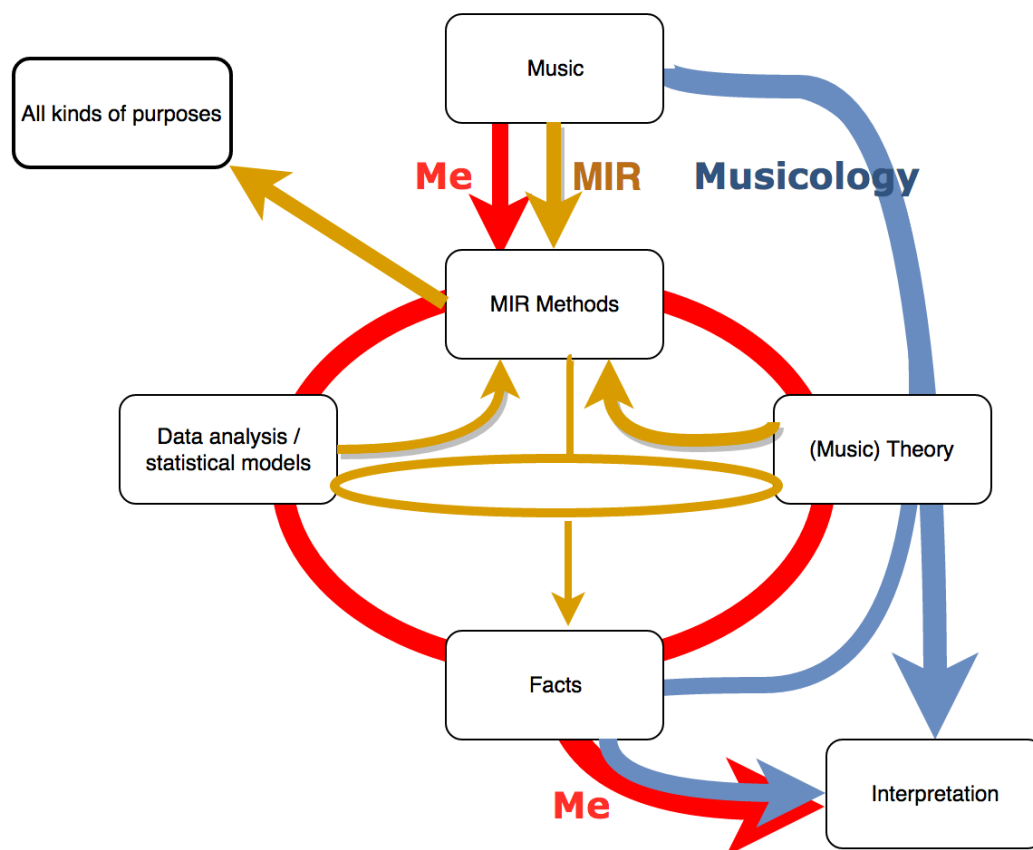


Figure 4 Diagram displaying differences in end goals between Musicology, MIR and me.

MIR and musicology are not only separate fields that have different objectives they also have their base in two distinct research traditions: Broadly speaking the computational or natural sciences and the humanities. Aucouturier and Bigand's fictive dialogue entitled *Mel Cepstrum e3 Ann Ova: The Difficult Dialog Between MIR and Music Cognition* (2012) reflects well some of the challenges establishing a fruitful dialogue between MIR and music cognition. Many of their concerns can be transferred to musicology more generally. The two fields

share a mutual interest in the same topic, human cognition of sound and music, but nevertheless, there are difficulties establishing a productive dialogue. The discrepancies covered in the dialogue relate to differences in culture, difficulties in translating from MIR metrics to natural language, questionable epistemological value of the MIR results, potential not-invented-here-bias, and ambiguities about what machine learning actually can learn us.

Grouped into two overall categories the problems covered in the dialogue relate to either differences that concern research culture, or to what we can actually learn or deduce from MIR features. The latter is most urgent to investigate in this thesis because we cannot learn to practice with new tools by talking culture alone. Nevertheless, the cultural issues play a role as well. If we have to collaborate better, we have to understand each other better, because “both disciplines suffer from this lack of mutual influence”, as Kranenburg et al. have asserted (2007, 3). The question about culture is important to consider because it can equip me with knowledge about how to translate between the two fields’ ways of communicating. In the dialogue, Ann Ova, the psychologist, for example explicitly disapproves MIR researcher's use of the word *semantic*. Alan Marsden has proposed that better interdisciplinary research could arise “from effort in understanding each other’s domain” and mutual “humility” (2012, 151). I will seek to pursue this.

3.2 Concern #1: How to Use Quantitative Methods in a Qualitative Discipline?

3.2.1 Quantitative methods in the humanities

“Scholars in the fields of the Humanities are habitually (and properly) afraid of statistical machinery” (Bronson 1949, 81)

This line is the first in the, to my best of knowledge, oldest musicological study that applies computational methods I know. It reflects well the cultural divide, and it is telling that this division of humanists into those who deal with numbers and statistics and those who do not is addressed almost as a premise. About 18 years later Raymond Erickson reported from the America Musicological Society in 1967 that there “would appear to be [...] a widening gulf between scholars who pursue the traditional methods of historical musicology and those who have adopted the computer as their chief research tool” (1968,

89). These historical examples exemplify well the general understanding of computation in the humanities.

In more recent times we find similar views: Kranenburg et al. for example in 2007 stated about the relationship between Musicology and MIR in general that “[a]lthough the topic is the same (music), there seems to be gaps in ways of understanding it.” While Burgoyne writes that one of the two major obstacles is that the two groups have “difficulty communicating with each other”. Because “[t]he needs and jargons of musicologists are alien to most music information scientists, who tend to originate from computer engineering, and engineers’ statistical models are opaque to most musicologists, who do normally not acquire sophisticated training in mathematics” (2012, 1).

One of the main points of dispute is what role quantitative data should play in research. The notion of proof is important here since there are rather different notions of proof in the respective cultures. This debate could seem to get to the heart of the matter of my research subject. For when analyzing MIR features, humanists will have to enter the quantitative domain, since large-scale analysis requires that you “perform some kind of quantitative analysis” (Wiering and Benetos 2013, 2, Tim Crawford paraphrased).

Most musicologists are not used to think about proof, at least in the same quantifiable sense as natural scientists. This is a point where the difference between cultures is expressed clearly. When researchers from data-oriented fields conduct large-scale music studies, the lack of evidence in humanities studies is often remarked. Serrá et al.⁴¹, for example, write that “many of these aspects remain formally unknown or lack scientific evidence, specially the latter”. And the statement is followed by the general comment that this “is very often neglected in music-related studies, from musicological analyses to technological applications” (2012, 1). Mauch et al. provide another example: “[M]ost claims about [popular music's] history are anecdotal rather than scientific in nature” (2015,1).

However, Cook and Clarke have detected that one problem with many objective music studies was “what one was meant to do with them, what their value was” (2004, 6). Later Cook concretized a little further that an empirical mindset often implies that the calculations become the final result:

41 The authors are from “Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC), Bellaterra, Barcelona, Spain – Complex, Systems Group, Centre de Recerca Matemàtica, Bellaterra, Barcelona, Spain, - Departament de Física Fonamental, Universitat de Barcelona, Barcelona, Spain - Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.”

“The problem with much musicological writing that adopts empirical methods is that it stop where the data stops, rather than using the data as a jumping-off point for the more informal and listening-based critical or historical interpretation that gives musicology its *raison d’être*” (2010, 12, italics in original)

During my project time, I engaged thoroughly with two MIR-assisted large-scale analyses of music (Serrà et al. 2012; Mauch et al. 2015), and I found that Cook’s concern remains the problem today: These analyses tend to stop where the data stops. It is also unclear what to do with them, mostly because they promise way more in the text than they can account for. However, as I shall demonstrate in Chapter 6, the value can perhaps be found somewhere else than in what is explicated.

The two positions regarding proof, which I have delineated above, were well demonstrated in a debate about quantitative methods in digital musicology. Wiering and Benetos have reported from the digital musicology session at ISMIR 2013:

“Matthias [Mauch, trained as a computer scientist], recently had his first experience on a humanities workshop, where there seems to be no notion on if something is true or not. Matthias asks if Tim [Crawford] intends to transform musicology as being more scientific. Tim responds that he indeed wants to change it, but not necessarily towards being more scientific. He explains that the notion of proof in humanistic research is completely different from scientific proof. For example, an elegant argument might be more convincing to a humanist than any amount of statistics. There is no objective truth, although there are many interesting ideas and influences.” (Wiering and Benetos 2013, 2)

3.2.2 How I relate to quantitative methods

The quote above is relevant for understanding how many musicologists relate to proof and consequently also how I approach it. Proof in the scientific sense is rarely the goal for the humanities scholar. Rather, with quantitative methods, it will be necessary to negotiate find a good balance and between between the quantitative and qualitative. Helle Porsdam explains:

“It is not a question of science/technology versus the arts and humanities - but instead a question of finding the right balance between quantifying and qualitative ways of thinking” (2013, 37).

Because of this wanting to balance ways of thinking my primary focus is not in measuring how well MIR measures quantifiable music analytical concepts, such as tempo or the mode of a song. Rather, I will investigate how MIR can be applied as a tool for musicological purposes. It is the interplay between the many components⁴² I mix that interests me. Understanding data is one element, but understanding how to gain music analytical insight by exploiting data is even more important. How can we establish connections between data, information, and interpretations, and back again? Or how can we create a process in which “observation leading to interpretation and interpretation in turn guiding observation” (Cook and Clarke 2004, 3). With observation in this case mostly done by machines and humans in collaboration.

Therefore, it is necessary to distinguish between quantitative ends and means. To think of only one way of applying audio content analysis would be to restrict this study too much. That is also why I above wrote that this debate about evidence only *seemed* to get to the heart of the matter. There are many other purposes one can use statistical methods for, and they may show to be equally relevant or contain as much value as others. In his bibliography about statistical application in musicology, Nigel Nettheim divides statistical methods into three categories, “descriptive, exploratory and confirmatory” and explainsthat “musical applications to date are primary descriptive” (1997, 1. introduction). That is, they summarize basic and simple information about the dataset. Nettheim wrote this in 1997, and I still see this to be the primary application of large datasets (for example demonstrated in Michel et al. 2011). However, today’s machine power adds a better opportunity for computers to be a practical tool suitable for “exploring” the dataset and “evaluating usefulness of concepts” (Kranenburg et al. 2007).

In my case studies in Chapter 5-7, I will apply the techniques for mainly explorative and descriptive purposes. As a tool for informing my knowledge about a dataset of files that are larger than I could manage manually and as a tool for assisting my listening. John Tukey’s data analytical prescriptions sum well up my approach to data analysis:

- “(b1) Data analysis must seek for scope and usefulness rather than security.

⁴² Music analysis, data mining, music information retrieval, acoustics, big data, information science, digital humanities’ and digital musicological theory, and the software I use.

- (b2) Data analysis must be willing to err moderately often in order that inadequate evidence shall more often suggest the right answer.
- (b3) Data analysis must use mathematical argument and mathematical results as bases for judgment rather than as bases for proof or stamps of validity” (1962, 6)

3.3 Concern #2: How to Interpret MIR Features?

3.3.1 We do not know how to translate from ACA to music analysis

Rather than teaching and discussing the role of confirmatory statistical analysis and p-values to musicologists, I regard it is as more urgent to investigate what MIR methods show in music analytic sense. Crudely put, how useful is it that we have proven something if we cannot explain what we have proven? The fictive dialogue mentioned above provides a fine example of this. Ann Ova, the psychologist, presents a diagram of values showing how well different ACA features correlate with values of valence and arousal. She tries to conclude from it:

“We see stimulus valence is very well explained by, let me get this right, the *entropy of the period of the magnitude of the maximum peak detected every 50ms in the signal's chromagram* (a chromagram, as you know, gives at each successive time position the energy observed at the frequency corresponding to each note - c, c#, d, etc., of each octave). Similarly, stimulus arousal seems to result from the *standard deviation of the 6th MFCC* and the *mean of the 3rd*, and - mind you - not the opposite. [...] See what I mean? That surely fits well to the data, but I'm sure you realize it does not actually *explain* anything.” (Aucouturier and Bigand 2012, 398, italics in original)

Honingh et al. have addressed this issue more generally. They explain “one of the problems [of applying MIR in a musicological setting] is, how to interpret features developed in MIR in musically meaningful ways [...], such that MIR research might contribute to musicological research” (2014,1). I regard this interpretation problem as a significant problem to solve before musicologists can deploy the methods for useful purposes.

One of the reasons why most digital humanities theories and studies I have encountered stem from a literary tradition is presumably because of a shorter way from digital text

analysis methods to their interpretation. Words are easy to count. With the aid of computers, we can count words in millions of digitized books reaching strong indications about the topics covered. Or about the cultural adoption of new technologies, the censoring of different artists in Germany 1933-1945, and the development in verbs' regularity and irregularity as Aiden & Michel have demonstrated (2011; 2013)⁴³. We can model topics of passages from the most frequent words in passages to find out what topics are the most common within a corpus (Jockers 2013). Or we can map relationships between characters in a novel, by counting the names' co-occurrences in passages (Moretti 2011). Just to name a few examples. And yes, similar operations are possible with music. One example is the Music Ngram Viewer⁴⁴ where you can trace the occurrences of any motivic pattern in the past 1000 years (Viro 2011). Nevertheless, it is my assertion here that when it comes to music interpretation becomes even more complex. And the interpretation only becomes even further complicated when analyzing audio.

For the MIR researcher lack of explanation or interpretation is often not a problem. The creation of a model that estimates emotions is an example of a MIR task. (This task would probably correspond with the block quote above). The MIR researcher would then want to create the best model to predict emotions from measuring audio files. And *best* depends on its precision: How well can the model predict the emotions in a piece of music if fed with an audio file? For such a task, the explanation of the relationship between musical features and emotions is not a part of the goal, and consequently not of importance. However, for the music psychologist who wants to explain the relationship between sound and emotion, this approach can become problematic. In the fictive dialogue, Ann Ova addresses the MIR researcher: "your discipline is interested in the result we are interested in the process" (Aucouturier and Bigand 2012, 401). This discrepancy regarding goals constitutes a general concern: MIR models are often constructed from an "engineering" approach, but this diminishes the explanatory value in the music analytic sense.⁴⁵

Returning once again to the adaptations of the sorts of judgments that, according to John Tukey (1962, 9), are likely to be involved in almost every instance of a data analysis:

- 1) the subject you analyse (the music)
- 2) how the data has been created, and what it stands for (MIR features)

⁴³ Or for the Ted Talk-version: https://www.ted.com/talks/what_we_learned_from_5_million_books

⁴⁴ www.peachnote.com, retrieved August 26, 2016

⁴⁵ This argument is inspired by Frans Wiering's presentation notes from "Balancing computational means and humanities ends in computational musicology" (Wiering 2012) Unfortunately, I did not attend the presentation, and I only have access to the powerpoint slides which are formulated as bullet points. Consequently, I am not entirely sure whether Frans Wiering would agree.

3) the particular data analysis techniques applied.

It seems obvious that 2) is the most apparent obstacle to musicologists. We lack knowledge about this source of judgment: Whether the musicologist wants to understand or conduct a data-driven analysis, or she or he wants to conduct his or her analysis of audio files, the understanding of MIR features and what they mean musically will in many cases be the primary obstacle.

3.3.2 Why understand what MIR features represent?

Referring again to the quote above, Ann Ova, the music psychologist, wondered what to conclude from the feature “entropy of the period of the magnitude of the maximum peak detected every 50ms in the signal's chromagram”. But the fact that she does not understand it does not necessarily imply that it is useless and cannot be translated into music analytical terms. Later in the dialogue, the two researchers agree that music psychologists might have a “not-invented-here bias”; and that music psychologists “can’t be bothered” (399) with MIR, because it looks complicated.

But there is no way around it. If musicologists have to exploit the potentials of ACA methods or understand what they infer, they will have to be “bothered”, at least to some extent. For my case, I want to conduct a music analysis deploying audio content analysis methods, and I found that the best piece of software that was accessible to me was MIRtoolbox. MIRtoolbox calculates many standard MIR features, often used by MIR, but largely unknown by musicology. Consequently, it would be bad practice not to engage with and examine the individual features. It is implicitly written in my research question to find out whether they should bother (more) about understanding ACA.

Additionally, other disciplines are conducting music research with ACA methods, and musicologists may want to tap into this dialogue, which is a point made by both Leman (2008) and Wallmark (2013). Leman writes from the viewpoint of systematic musicology, polemically he asks “who stole systematic musicology?” However, he reformulates the same question into a positive one: “What is of such a value in systematic musicology that it can appeal to a broad range of researchers working in other disciplines?” (92). Leman focuses on how musicological knowledge can contribute to transdisciplinary music research, arguing how MIR can benefit from musicology. Zachary Wallmark discusses the relationship from the other way around, how MIR can benefit to musicology. Thereby this text is closer related to my topic. He, like Leman, has observed that ““scientific” studies

about music are [...] likely to travel quickly through the mediasphere, broadly influencing the public's understanding of music in venues rarely accorded to musicology" (4).

Wallmark addresses musicologists: "[d]o we really want a bunch of engineers and computer scientists setting the public agenda for music scholarship?" (4) And he argues for an expansion of the "methodological toolbox", otherwise, he argues, "we're going to be left out this important and necessary conversation" (4).

Another good reason why it can be a good idea to understand what MIR features represents is that it can enable us to deduce knowledge about music from MIR experiments. Though this is slightly off my topic. Honingh et al. have argued "that it often remains unclear what has been learned with machine-learning approaches" (2014, 1) (see also W. Bas de Haas and Wiering 2010). By engaging with MIR's way of analyzing, I hope to, as a side effect, improve the understanding of the underlying mechanisms in MIR results. Thereby musicology will be better equipped to deduce knowledge from machine-learned tasks, and to be better able to criticize them and use them themselves for analytical purposes. However, due to the engineering approach, the statistical calculations behind the models are often extremely complicated.

3.4 How to Solve These Concerns

Below is a diagram (Figure 5) of the components of my object of my interest. Within the blue circle are the components in a prototypical music analytical application of MIR for large-scale studies. This way of analyzing many audio files at once is consequently also my primary focus. However, the individual topics within the boxes are covered more extensively elsewhere and will therefore not be in the main focus of this thesis. Rather, I will focus my attention at the arrows and the process in its entirety.

The diagram at the same time displays where I have chosen to delimit my focus of attention. I will reflect upon how we can start with MIR features derived from a corpus of audio files (1), apply a combination of data analysis and music analytic methods on the created metadata (2), and arrive at conclusions about the musical and acoustic characteristics of the corpus (3). In other words, I will only go as far as I feel that the measurements and listening allow me to. This focus at the same time explains why more general theories about music's context such as gender, social construction, body, etc. will not be pursued here. The whole process will be in line with Helle Porsdam's advice that

"[t]he many new developments within DH must be discussed with a view not only to their potential, but also to their limits" (2013, 42).

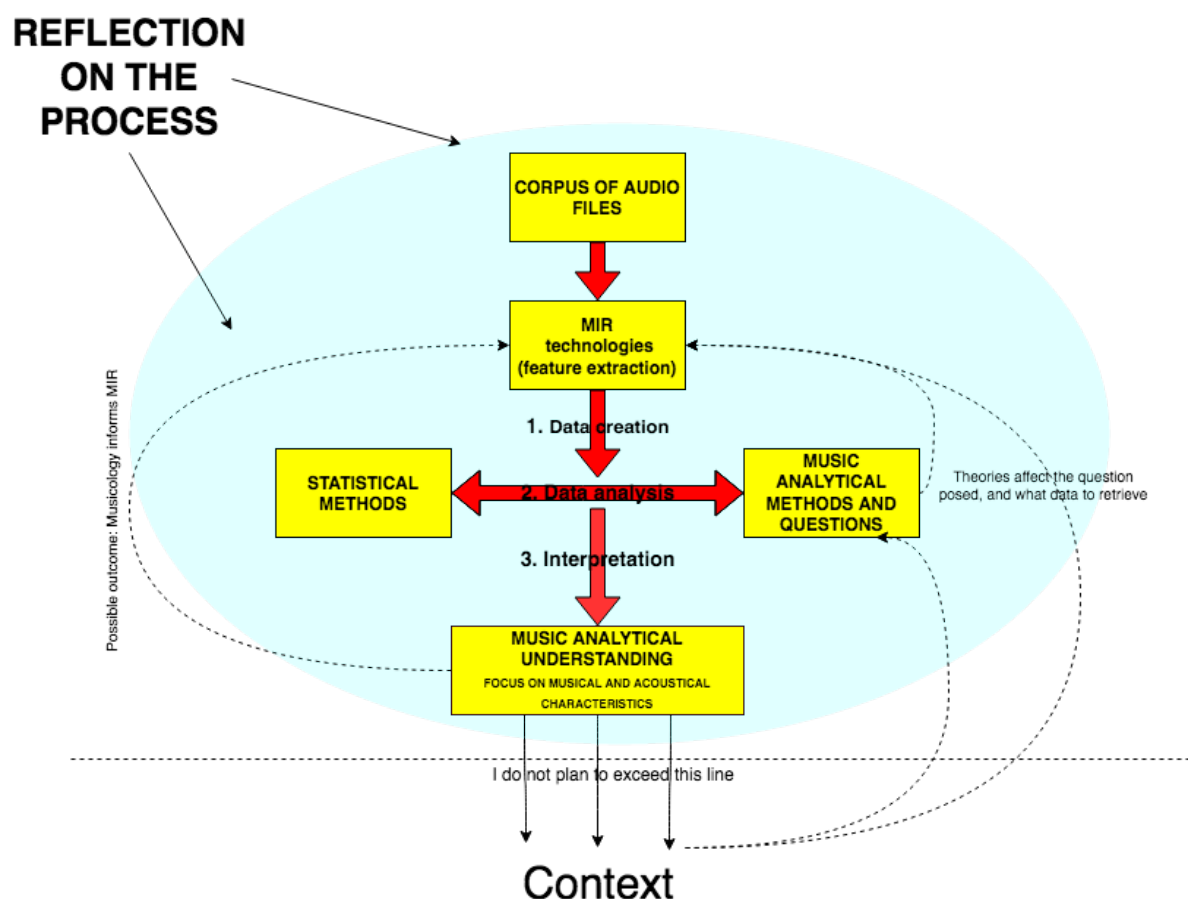


Figure 5 The research components of my interest. The blue circle is the flowchart of a prototypical MIR assisted large-scale analysis.

There are two sides of data creation (1), one that relates to how ACA methods create the data, and one that relates to the management of this data. Especially the first point affects (3) because what we can understand from an analysis is, of course, closely connected with what is measured. For me, the data analysis process (2) will be a rather practical task, an interplay between data analysis and music analysis informed by knowledge about the fields I combine. In relation to (3), I have argued above in 3.3 that the gap between musicology and MIR especially manifest itself in the absence of (3). One prominent assumption here is that we need to examine and experience through practice how to progress from sheer data analysis towards interpretation. This point is not covered very well and establishing a better linke is very beneficial for my target group, musicologists. Hence, one of the goals here is to contribute to creating an “understanding of the relationship between facts and

interpretations” by “not reducing interpretations to facts” as Cook and Clarke saw as the goal of empirical musicology (2004, 3). I will pursue knowledge about step 3 through what I will refer to as *methodology-informed* listening. This mode of listening seeks to combine my own listening and music analytic knowledge with knowledge on how the algorithms calculate their features. The knowledge about the algorithms will direct my attention towards what musical aspects to listen for.

On not transgressing the dotted line

Most musicologists will have the need for relating the empirical information that ACA methods can provide to some musical context. So, exceeding the dotted line also relates to my research question according to which I search for the relevance of conducting large-scale analysis for musicological purposes. However, if I opened up for the wealth of possible theories that can be connected to large-scale studies, it would blur the focus of thesis. So instead I chose to illuminate the methodological toolkit that MIR enables.

3.5 Choice of Analyses

I will cover these issues theoretically in Chapter 4, while Chapter 5-7 will seek to illuminate them through empirical investigation. The three case studies, I have chosen, each illuminates different, relevant aspects of MIR in musicology. These cases cover some of the most important and critical aspects of my research question: They illuminate the cultural divide, they exploit and investigate potential in both low- and high-level metrics, and they practice large-scale analysis with the tools.

Case 1) Echo Nest's features

In chapter 5, I will investigate Echo Nest's metrics. They are an example of how digital methods can model and thereby create new and more intuitively understandable music features. This case has been chosen to inform the discussion of high-level MIR features, which according to Burgoyne, Fujinaga, and Downie “would be of greater musicological and cultural interest” than low-level features (2015, 222). I will, in this case, focus on step 1, data creation, and how it relates to step 3.

Case 2) Mauch analysis

In Chapter 6, I will examine an already conducted large-scale ACA assisted study (Mauch et al. 2015) This study is an example of an analysis that has gained a lot of attraction in the media (as Wallmark 2013 describes). It is also an excellent example of the potentials of

ACA, and consequently, it provides inspiration on how to think with ACA tools. However, at the same time, it also serves as an example of the above-discussed cultural divide between the sciences and humanities. The epistemological value of the analysis is somewhat opaque, amongst other due to a complicated relationship between data analysis and music analytic value. It is an example of an analysis that practically stops at step 2 in my diagram above, and my analysis will be a discussion on how to progress to step 3 from it. Thus the main purpose of choosing this case is to equip musicological scholars to better understand ACA assisted analyses, the complex algorithms behind them, and the cultural divide.

Case 3) Analysis of DJ sets

Case 3 will be the main case study, and it will cover all the steps in Figure 5. I will demonstrate a real application of the tools by conducting a corpus study of 89 DJ sets played at the Electronic Dance Music Festival, Ultra in Miami, 2015. This corpus study has two overarching purposes: Firstly, it has a music analytic objective. Thereby it stresses the *musicological purpose*, explicated in my research question. Secondly, for the analysis itself to be a help in answering this question, it should illuminate methodological aspects. Consequently, the analysis was both constructed and conducted to get hands-on experience applying MIR technologies; to maximize the reflection and evaluation on the techniques' usefulness for creating knowledge about western popular music. Hence, this study is an ingredient in a *show, don't tell* strategy for answering my research question: I will practice digital musicology, because of the shortfall of musicological studies applying ACA methods, in comparison with the large amount of methodological reflections. Practicing digital musicology seeks to shed a music analytic light on theories, and it seeks to gain hands-on-experience which is crucial for getting accustomed to the caveats and pitfalls the methods entail. Hopefully, this study can contribute to the development of best practices on how to understand MIR values.

3.6 Rounding off Chapter 3

In this chapter, I have accounted for my focal points in the pursuit of answering my research question. I identified that there especially was a need for being able to understand MIR features to progress with ACA assisted analysis for it to become more than a sheer data analysis. This step is necessary to clarify before we know how ACA techniques can inform music analysis. Thereby, Schnapp and Presner's distinction between digital

humanities' waves becomes useful to explain my position and ideals. They identified two waves of the digital humanities: The first wave was:

“quantitative, mobilizing the search and retrieval powers of the database, automating corpus linguistics, stacking hypercards into critical arrays. While second wave is qualitative, interpretive, experiential, emotive, generative. It harnesses digital toolkits in the service of the humanities core methodological strengths: attention to complexity, medium specificity, historical context, analytical depth, critique and interpretation” (2).

While it was the CoSound engineers' task to take the project into the first wave, it became important for me to take my project into the second. I wanted to close the gap, by contributing with my “core methodological strengths,” complying with my target group's requests.

In this chapter, I also touched upon the cultural divide between MIR and musicology. This divide pervades the current relationship and affects the modes of thinking within the respective disciplines and about each other. This cultural divide is, however, not my main focus. Rather it is in the back of my head throughout the thesis. It is about finding “the right balance between quantifying and qualitative ways of thinking”, as Helle Porsdam has explained. (2013, 37). She continues, “[b]oth are important - and both offer us something that we cannot do without”. Who says that we necessarily lose something on one side, if we put our emphasis on the other? But to judge where the right balance is, or whether we can create synergies, we must know more about the tools, what they measure, and the new types of questions they enable.

CHAPTER 4

From Audio Content Analysis to Music Analysis

Where Chapter 5-7 contains my case studies, and thus serves as the empirical investigation of ACA methods' usability, this Chapter 4 has the purpose of providing a theoretical understanding the potentials of the methods, and how to practice music analysis with them. The two most important questions that I pose in this chapter are:

- *What kind of data can ACA create?* In this part, I will briefly introduce to MIR, and audio content analysis. The purpose is to improve the understanding of the features, and I will pursue this by delineating the basic thoughts and calculations that underlie the tools.

- *What are prospects of creating a lot of data for music analysis?* In this part, I will illuminate how data creation, and more specifically ACA methods, can be incorporated into musicological practices. I will discuss what to gain from this, and which points that require special attention.

4.1. Data Creation with ACA

As I explained in Chapter 2, MIR and its methods are covered comprehensively elsewhere (Meinard Müller 2015; Li, Ogiwara, and Tzanetakis 2012; Lerch 2012). While I, in Chapter 3, argued that my main focus would not be on the individual components, I mix, but rather on how to integrate them and create synergies. But obviously, it is hard to illuminate how to apply one thing within another without knowledge about both. To cross from one thing to another depends on where you come from, and where you are heading.

Many musicologists know only little about MIR and ACA, and therefore, I will in this part of the chapter illuminate the basics of ACA, with particular attention to what I find most music analytical relevant.

ACA methods are very complex algorithms calculating aspects from digitized audio, which consists long strings of 0's and 1's. This makes the methods rather opaque and impossible to comprehend entirely. Ideally, there is a transparent relationship between methods and measurements, but with ACA methods this ideal seems almost impossible to achieve. Although the methods are formalized and consistently applied they are not necessarily transparent, quite the contrary. (I will exemplify this in Chapter 5 and 6.) As a consequent, it is often also hard to determine why the results turn out how they do.

I have chosen to approach ACA by acknowledging this gap between understanding the methods entirely and sufficiently enough for being able to practice with them. This is most likely a situation that musicologists will have to get to used to, if wanting to exploit the powers of the tools. For my case, if I had unlimited time at my disposal, it would, of course, seek to understand the algorithms in details, especially those I apply for analysis. But I do not have that amount of time. Nobody has. But this does not exclude the tools from being useful. Moreover, if I emphasized on detailed understanding, this piece of research would quicker become obsolete, because the methods occasionally change when a better algorithm for a certain task is developed. Instead of scrutinizing the individual algorithms, I will, therefore, in this chapter focus on disseminating general principles that underlie ACA methods, and discuss how it can inform music analysis. This approach fits one of the goals in this chapter, which is to stimulate how to think with ACA methods.

4.1.1 What is MIR?

For my purpose, Stephen Downie's definition of MIR is well-suitable:

“Music Information Retrieval (MIR) is the process of searching for, and finding, music objects, or parts of music objects, via a query framed musically and/or in musical terms.” (CIRMMTvideo 2012, 14:00)

This definition elucidates the difference in purpose of MIR compared to musicology's. In the presentation, Downie explains that MIR aims at finding “things”, such as “scores, parts, recordings” (2012, 14:00). For my purpose, it is equally relevant that to improve search precisions better methods are desirable. It is these methods, the “musical terms”,

which MIR uses for searching that are of interest here. Downie explains that the musical terms, for example, can be “genre, style, tempo, etc.” (14:00). The etc. opens for a wealth of opportunities.

These musical terms can be derived from different sources. Within MIR there are “at least three views on musical data coexist; that is, music as represented by metadata (originating from the library science subcommunity), by encoded notation (from musicology) and by digital audio (from digital signal processing)” (de Haas 2012, 3). This thesis pivots around the third type of data: Digital audio as musical data, derived from digital signal processing techniques. And therefore it is sufficient to note that any measurable aspect of an audio file principally can become a musical term.

4.1.2 What does ACA measure?

Digital signal processing on audio has its origin in speech processing, but the techniques have been found useful for music too (George Tzanetakis 2014; CIRMMTvideo 2012). As a rule of thumb, we can derive rhythmical information from measuring aspects of the strength of the signal (George Tzanetakis 2014; Lartillot 2014). Measuring when accentuations occur can, for example, form a basic approach for estimating tempo. We can derive timbral and tonal information by applying the Fourier transform (FFT), which has been called the most important mathematical operation in audio processing (Müller 2015, 69). The basic mathematic idea behind the FFT is that any waveform can be summarized as a sum of sinusoids. Consequently, you can apply the FFT on a waveform and calculate information about the frequency content; how much each frequency is present in the music at the given time.

It takes some mathematical operations to transform the sound signal into useful music analytic measures. In this process, a number of psycho-acoustical aspects have to be built into the model (M. Müller et al. 2011)⁴⁶. Our perception of harmonic content is, for example, logarithmic: The note C in one octave is harmonically related to C's in other octaves, but the distances between the C's are not equally, equidistantly spaced. Rather, the distance doubles for every octave. However, we can calculate such adaptation into a model that calculates tonal content. The general message is that we, by applying mathematical operations to audio files, can calculate a wealth of information about music. ACA can, for example, estimate traditional, well-known musicological features such as

⁴⁶ The text also accounts for many of the concrete challenges that occur in the process.

tempo (Scheirer 1998), chroma (pitch content) (Wakefield 1999) and chords (McVicar et al. 2014) somewhat successfully.

Besides, new born-digital features have been enabled as a consequence of digital music and digital methods. *RMS energy*, *Spectral Centroid*, *MFCC's* (roughly translates to timbre) and *Zerocrossrate* are examples of these. I will investigate these and other new feature in my analysis of 89 DJ sets, in Chapter 7. A general principle for applying MIR features is that, as Downie states, “there is no a priori theory that says you need feature X” (CIRMMTvideo 2012, 22:00). Therefore you can in principle apply any thinkable MIR feature for your analysis. ACA allows high flexibility in measurements. So it is, in fact, possible to measure thousands of features from each audio file. All sorts of mathematical operations are possible, you can combine features, or alter the mathematical operations to create new features that fit the given purpose best, or that might be more music analytic relevant.

Machine learning is an important principle within MIR, and machine-learned features are examples of combining features for higher-level purposes. By calculating many features from a bunch of annotated audio files, the computer can recognize patterns in how the features relate to the annotations. This technique have been deployed for classification tasks, such as genre classification (G. Tzanetakis and Cook 2002), estimation of mood (Laurier, Grivolla, and Herrera 2008), composer detection (Herremans, Martens, and Sörensen 2014a), instrument (Essid, Richard, and David 2006), music similarity (Schnitzer, Flexer, and Widmer 2009), and many others. Machine learning techniques also allow for calculating new and more intuitive music features. The company, Echo Nest, apply them amongst others for automatically estimating values for songs' valence, danceability, acousticness, etc. I will discuss the prospects and limits of machine learning in section 4.7 of this chapter. And I will investigate the musicological usability of Echo Nest's machine-learned features in Chapter 5.

In appendix 1, I have provided a list of MIR resources, such as datasets and software.

4.1.3 The link to musicology

ACA methods allow a high level of flexibility. It allows us to automatically create vast amounts of musical data about any audio file: We can measure acoustical aspects from time units smaller than humanly perceivable, we can summarize this information to perceptible time-levels, or summarize for whole songs. In principle, we can measure attributes of

corpora larger than it will ever be possible to listen to for a human being in an entire life. We do not have to rely on human annotating the music and we do not have to rely on scores.

But we should also keep in mind that we create data, not because of the data itself, but because data can serve a larger purpose. Christoph Schöch has explained very precisely that data “is not the object of study, but “stands in” for it in some way” (2013). This principle of data representing something else is not unfamiliar to most musicologists. A music analysis of scores is often an analysis of the music, and not of the scores. Thus scores fit into Schöch’s statement about data; they are also representations of the object of study, the music. Nicholas Cook argues similarly:

“It is true that a Schenkerian analysis, say, looks as if it were an analysis of scores. But in fact it is not. Rather it is using the score as a convenient, and tolerably adequate, way of talking about the real topic of musical analysis, which is the analyst’s (and hopefully the reader’s) experience of the music.” (Cook 1987, 228—emphasis in the original)

I take a similar standpoint about data: Data is an instrument that can assist other purposes. But in relation to the Cook quote above, I want to clarify what I also wrote in section 2.2, that the purpose of analyzing many pieces of music will often differ from the purpose of analyzing an individual piece. The analysis of many pieces of music will, for example, not necessarily illuminate the experience of the music.

4.2 The Advantages

In the following pages, I will, firstly, present the situation from a level of potential. I will, from a theoretic viewpoint, elaborate on the new possibilities that arise as a result of digital techniques combined with large amounts of quantitative data. Next, I will discuss how to apply the data analysis techniques for humanities’ purposes. Finally, I will address the question of whether we can trust ACA techniques, and what questions we ought to ask them.

4.2.1 The benefits of digital methods

Though datafying music is not the end-goal, it enables us to create and process measurements of music with computers. And computers that bring along a lot of advantages: Computers are good at performing the same task over and over again, and they can do so multiple times quicker than humans. These strengths are the primary reasons why computers are attractive for conducting large-scale analysis of music. They enable us to measure many aspects about many songs consistently, and thereby they can create huge amounts of ACA data. Thus we can enlarge the scope of our analysis multiple times. Before digital tools we were practically limited to investigating only a few instances, now we can investigate many. And the speed of such operations only increases.

One advantage that comes with the digitization of vast amounts of cultural objects is that we do no longer necessarily have to rely on sampling. We now can investigate whole corpora (Huron 2013; Manovich 2012, 250; Mayer-Schönberger and Cukier 2013, 10). One of the changes that arise with computers' ability to analyze all is that we now can apply computers to help us overcome the limitations of canons. Traditionally, scholars tend to read only a fraction of all works, but one problem is how representative these are? Digital methods can help overcome this limitation and help us "read" whole corpora of texts, which can give us a wider, and more comprehensive understanding of styles, periods, artists, etc. (Moretti 2005; Wilkens 2012). Musicology is now facing similar conditions.

4.2.2 The value of statistics

When we want to analyze large datasets, we have to perform some kind of quantitative analysis. Statistics are very good at zooming out; they can give us a view of the bigger picture, create compact descriptions of entities, crude descriptions of the world, identify outliers and a lot more. But one price is that we lose sight of the detail when we zoom out. Literary scholar, Franco Moretti, has called this zooming out distant reading and describes it as a "condition of knowledge":

"If we want to understand the system in its entirety, we must accept losing something. We always pay a price for theoretical knowledge: reality is infinitely rich; concepts are abstract, are poor. But it's precisely this 'poverty' that makes it possible to handle them, and therefore to know. This is why less is actually more."
(2000, 58)

Matthew Jockers, for example, have modeled “topics” from automatically counting the most frequent words in a text (2013, 118-153). By detecting which words often occur close to each other, topic modeling is a useful way of getting overall impressions of what a passage is about. For example, the words *Indian, chief, warrior, men, party, knife, stream, tradition, mountains* occur often near each other in the examined texts (2013, 126). This technique of forming topics out of large clusters of words enables us to grasp overall structures in the complex material of thousands of book pages. It allows us to zoom out from the single text and examine more general questions about what topics were covered at a given time in history. One could imagine similar musicological questions to be studied this way. Mauch et al. (2015) approach music analysis a similar way. They create timbral and tonal “topics” that allow them to manage their corpus of 17,000 songs, making overall trends graspable.

In the quote above, Moretti argued that what we lose in detail we win in the ability to see larger patterns. However, it is not necessarily a question of either or. It is not necessarily only a question of choosing whether to ACA methods and understanding the entirety or applying traditional hermeneutic reading and understanding the complexities. The data approach allows us also to nuance generalizations (Cook 2010). We can use statistics to examine distributions, and we may, for example, find that history may change less radically, or abruptly than it sometimes can seem retrospectively; when observed through the canon. The graphs in Figure 2 in Mauch et al. (2015) are examples that indicate tendencies of the speed of the change of popular music in the USA.

4.2.3 An inwards expansion of the perception

The data view not only allows us to zoom out either. It also allows us to zoom in on the individual. Lev Manovich proposes the notion of “*surface data* about lots of people and *deep data* about the few individuals” (2011, 2, I exchanged quotation marks with italics).

Translated into a music analytic context, we can retrieve data regarding an immense amount of aspects also about time units even smaller than a human can perceive. Thereby we can for example register subtle differences that we perhaps do not perceive with our ears. ACA thus enables an “inwards” augmentation of our perception, and consequently, it allows an expansion of the level of detail in comparison to traditional music analysis. For example, when applying the MIRtoolbox (Lartillot 2008) for measuring *mirzerocross*, “the number of times the signal crosses the X-axis” (Lartillot 2014, 123). Or when calculating *mirinharmonicity*, “the amount of energy outside the ideal harmonic series” (2014, 143). In both cases, the music’s constituent elements which traditional music analysis investigates

are calculated on levels imperceptible to the human ear; on microscopic temporal and timbral levels. Nevertheless, they measure aspects of what constitutes the music's sound. And both measures can be associated with humans' perception of sound. Lartillot states, for example, that *mirzerocross* is an indicator of noisiness (2014, 123). I view *mirinharmonicity* as an indicator of dissonance.

4.2.4 New questions becomes enabled

Matthew Jockers provides a highly inspirational list of new questions about literature that we now can retrieve insight about with digital methods (2013, 27–28). Most of them can be translated into a musicological context. Freely translated they amongst other concern:

- patterns employed over time, across periods within regions, or within demographic groups
- cultural and societal forces that impact style and the evolution of style
- the waxing and waning of sound idioms
- the tastes and preferences of the establishment and whether and how they correspond to general tastes and preferences

Musicologist Arthur Mendel wrote in 1969 a similar list containing questions about 1600th-century music (52). His organization of the questions into either technical (statistical properties of notes in the scores), aesthetic (which regards modes of expression), and reflections of influences from outside music is interesting in this context. Mendel does not explicit remark it, but his categorization demonstrates that there are various levels of complexity between the types of questions: The broader the question, the less focused on the music itself, and the more complex it is to account for quantitatively; it will take either an increasing amount of formalization, more advanced statistics or increasing amount of interpretation.

Most of Jockers' questions would be categorized under Mendel's two latter categories. I am not sufficiently qualified to account for what concerns arise when trying to answer these questions with automatic text analysis tools, but when it comes to music a lot of new complications arise. They relate to the complexity of music. When for example trying to answer the question "whether successful works of [music] inspire schools or traditions" (Jockers 2013, 28), it gets very complicated. We have to somehow choose between a vague, opaque general answer and a less extensive, exhaustive, but more specific and precise answer. On the one side of the generality-precision spectrum, we could approach

the question by setting up a similarity metric⁴⁷. This is possible to some extent, as I will demonstrate in Chapter 7 (see also Collins 2012). But as Esparza, Bello, and Humphrey explain there are a lot of approaches to musical similarity, but none of them are really comprehensive: “[T]here is no unified view on what ‘similarity’ means, the shared intuition agrees that it is a multidimensional concept, incorporating a variety of socio-cultural and musical aspects” (2015, 39). On the other side of the spectrum, we could split the question into components: to what degree has the respectively rhythmical, timbral and tonal properties of successful work X inspired tradition Y? Or what characteristics of work X have been persevered in tradition Y? In any case, both approaches require that we formalize our question into quantitative statements. I will return to the question of formalization of methods below in Section 4.4.3.

4.2.5 Why we should do it?

Matthew Jockers proclaims that “[t]he goal of science, we hope, is to develop the best possible explanation for a phenomenon” (2013, 5). And to pursue this goal one has to apply the methods that fit best, he argues. In theory, ACA methods can potentially improve disciplines within musicology on at least three levels: 1) We can pose new questions that were practically impossible to pursue priorly. 2) We can create a firmer basis for our knowledge. 3) It can guide or even modify our thinking.

- 1) I have already argued for this point. And I will seek to exemplify throughout thesis.
- 2) By helping us examine the many, digital techniques can help us create more reliable and complete background information for our studies. Nicholas Cook and Eric Clarke have seen room for improvement in current musicological practices. They assert that:

“musicologists frequently work with very small amounts of data even where large data sets are available, resulting in findings that are less firmly grounded than they might be” (2004, 3).

While Huron has focused on that amount of evidence gathered is a way of enhancing scholarship:

“In nearly every case, scholarship is enhanced by the availability of additional evidence. Like prosecuting attorneys, scholars have a moral obligation to seek out

⁴⁷ I apply the word *metric* in the statistical sense. See Section 1.7.

additional sources of evidence/data whenever these can be obtained” (D. Huron 1999, 186).

But, as I explained in Chapter 3, evidence does not necessarily have to be understood in the scientific, natural science sense of the word. Nicholas Cook (2005) instead focuses on the tools’ potentials for posing new questions and for providing an extra add-on for musicology: “working with fuller data and larger data sets can open up new areas of musicology, but it can also mean doing traditionally musicology better.”⁴⁸

Correspondingly, Jockers never applies the macroanalysis as a final stamp, but always as a jumping-off point for discussion. Thereby the methods work as tools that can “enhance the perception” (Foster 2011, 19) because the digital methods enable him to view his corpus in ways that were not possible without them. It would not be practically possible to grasp a dataset of 106 nineteenth-century novels manually as well as with digital tools.

David Huron agrees on this point:

“[Q]uantitative methods are important for the same reason that musical notation can be important: like musical notation, quantitative methods allow us to observe patterns of organization that might otherwise be difficult or impossible to decipher. [...] It has everything to do with becoming a more observant music scholar. (1999, 155)

Digital methods can thus act as a way to observe things that would otherwise be very hard to see. The results may either confirm our expectations and thereby strengthen existing theories or they may reject them, and thus they challenge what we thought we knew (McCarty 2007). They thereby can help us look at our corpus in new ways, as a method that can help us shut our strong eye⁴⁹. Potentially it can help us overcome “the tendency to hear what you expect to hear” (Cook 2010, 12–13).

⁴⁸ Noteworthy, there is a similar vagueness towards the notion of proof in other humanities’ big data thinkers. Manovich, for example, neither mentions proof nor evidence but instead he more generally remarks that big data techniques “can allow us to make more confident statements about the field at large.” (2012, 252)

⁴⁹ Credit to Tanya Clement for this metaphor, which I heard at the Digital Humanities Conference in Lausanne, 2014.

4.3 How Humanities Objectives Fit with Data Approaches

But how can these advantages be achieved and exploited within the humanities? How can musicological scholars fulfill these potentials with assistance from the tools without dismissing their core competencies? And how can they incorporate the tools into their research practices and routines? I will theoretically elaborate on these questions in the remainder of this chapter. My main argument is that there are many ways to bring these advantages into humanities research, and there are multiple modes of collaboration between machines and human. But they all have in common that domain knowledge should never be thrown overboard.

Before directly engaging with the questions I proposed above, I will in this section set up some premises for the discussion. I see three main points that might appear as a conflict between natural sciences and humanities. They need clarification. The first is about thinking with big data, the second relates to evidence, and the third to humans' versus machines' listening. My main point throughout is that even though the methods are created within fields with certain ways of ideals and ways of practicing science, the same practices and ideals do not necessarily have to be transferred into the humanities.

4.3.1 Data answers *what*-questions, but humanists tend to ask *why* or *how*

Regarding the question why we should work with large amounts of data, Chris Anderson's text "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (2008) takes the position of the one extreme. Anderson claims that with enough data we no longer need theories or models (by which I believe he means theoretical models). "With enough data, the numbers speak for themselves," and "[c]orrelation is enough" are examples of big data enthusiastic catchphrases. Mayer-Schönberger and Cukier are almost equally as optimistic about the data deluge as Anderson. They similarly attribute a lot of power to the data itself: "We don't always need to know the cause of a phenomenon; rather, we can let data speak for itself" (2013, 14). They also make clear that "[b]ig data is about what, not why" (14). And yes, I will agree that a lot of data combined with data-driven approaches can offer interesting and new perspectives. It does also hold true to some extent that non-quantitative theoretical models become less important when you are creating quantitative models. This is the case for a lot of MIR activities. It also may be true that you can get a

long way with large amounts of data when you want to ask quantitative questions of e.g. what, where and when.

But in general, big datasets are useful for searching for quantitative correlations. And correlations do not necessarily imply causalities. This implies that there are a lot of other types of questions which are not included in Anderson's and Mayer-Schönberger and Cukier's ways of posing questions and these are the questions that musicologists most often will find most interesting. These include the qualitative questions that arise from the numbers; how to interpret the numbers and the methods and how to put them in context. More generally, humanists tend to pose mostly *how* and *why* questions. Often they are most interested in causes or implications of a phenomenon. My research question is an example of a *how*-question. I am interested in implications.

However, that humanists tend to pose qualitative questions does not necessarily imply that data analysis should be left out of any investigation of causes. Rather it is inevitable not to include empiri in musicological research, as Cook and Clarke argue:

“There is no such thing as a truly non-empirical musicology; what is at issue is the extent to which musicological discourse is grounded on empirical observation, and conversely the extent to which observation is regulated by discourse” (2004, 3).

Large-scale quantitative studies can play a similar role. They can work as empiri, which can become an element in the investigation of many other phenomena.

4.3.2 On evidence - embracing parsimony

As I argued in Chapter 3, there are also differences between the two cultures regarding how they think of evidence. Machines are usually associated with objectivity in the evidence-seeking way. However, as David Huron has explained, humanists tend to prefer “pluralism to parsimony” and “open accounts rather than closed explanations” (1999, 187). One might perhaps think that the increasing data-richness would have entailed ideals closer to the natural sciences, as Huron imagined, but not much seems to have changed in that regard. Neither in the humanities in general nor in musicology: Burgoyne et al. have for example more recently stated that “[m]usicological questions are more open-ended and descriptive than MIR questions.” (2015, 214). Musicologists still tend to have other incentives than to show if something is true or not.

Therefore it is pertinent to address here that the usage of machines does not necessarily imply objectivity in the rigorous, end-of-discussion sense. Jockers argue that rather than producing final statements the “macroanalytic approach simply provides an alternative method for accessing texts and simply another way of harvesting facts from and around texts. [...] The computer is a tool that assists the identification and compilation of evidence.” (2013, 29) The computer can, for example, become a tool-of-assistance that can inform other types of questions.

Just to make it clear; facts are not the problem *per se*. Rather it is the treatment and presentation of them that, according to Jockers, “render them “uninteresting” to scholars conditioned to reject the idea of a closed argument” (30). Hence, the prototype of a humanities scholar. The problem is that humanities scholars may find that facts presented as findings tend to close discussions, while interpretations tend to open them or suggest new modes of seeing the world. Jockers warns against “quantitative arrogance” of decisiveness, and he suggests that this perhaps has contributed to a general hesitation amongst literary scholars when it comes to the usefulness of quantification.

4.3.3 Humans and machines strengthening each other

The third point of seemingly conflict is machines’ objective versus humans’ hermeneutic reading. On the machine side of the spectrum ‘reading’ “implies a model that eschews human interpretation for algorithms employing a minimum of assumptions about what results will prove interesting or important”, as Katherine Hayles explains (2012, 47). Hayles thus in one line manages to bring up two components that traditional often have been strongly connected: The scientific ideal of objectivity because “human interpretation misleads”, and machines’ capability of being objective, since they are consistent, rigorous and enable a view from nowhere. To a large extent, this latter ideal also pervades MIR, and I will demonstrate it further in Chapter 6. In contrast, humanities scholars traditionally apply close, hermeneutic reading that “connotes sophisticated interpretations achieved through long years of scholarly study and immersion in primary texts” (Hayles 2012, 47) (see also Clement 2012, 883; Jockers 2013). For today’s musicological scholars the equivalent holds true: They manually listen to music, read scores and related books.

However, Hayles downplays the conflict: “[T]he tension between algorithmic analysis and hermeneutic close reading should not be overstated” (48). There will always be some degree of human interpretation in any process involving the computer, “for humans create the programs, implement them, and interpret the results” (47). Instead we should keep in

mind that machines and statistics can enable us new viewpoints of our objects of study. This is for example what Franco Moretti seeks to demonstrate with the concept of “distant reading” which emphasizes that machines can help us overcome our limited mental capacity:

“Distant reading: Where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes — or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more.” (Moretti 2000, 58)

This distant view does not imply objectivity or eschew human interpretation per se. As Hayles argues about distant reading, “[h]uman interpretation remains primary but is nevertheless wrenched out of its customary grooves by the scale at which ‘distant reading’ occurs” (47). When reading distantly, you interpret your objects of analysis at other levels.

4.3.4 Summing up on humanities and data

In this Section 4.3, I have argued that humanities scholars may have other incentives than many of the traditionally data-rich fields and the fields that produce the tools. But yet, there are many viable ways of applying digital methods for humanities’ purposes. Ways that on the one hand stir up traditional oppositions of conducting research and which at the same time can enable elements from both cultures to be brought into play. A similar way of thinking with the tools could be transferred into musicology, for computational musicology and musicology to integrate better. Musicologists need not necessarily link large-scale empirical inquiry with objectivity in the positivistic sense. If they want to apply data analytic approaches, they should focus on understanding the relationship between facts and interpretations better.

Or to quote Charles Darwin:

“About thirty years ago there was much talk that geologists ought only to observe and not theorize; and I well remember someone saying that at this rate a man might as well go into a gravel-pit and count the pebbles and describe the colours. How

odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service!”⁵⁰

Today, musicology could face a similar dilemma: ACA techniques make it very easy and fast to count a lot of aspects for us, but we have to find out what we want to use all this data for.

4.4 But What to Learn about Music with ACA Methods?

4.4.1 The value of facts

With today’s technological progress, the Darwin quote is indeed relevant. We can produce endless arrays of numbers measuring aspects of music. If research had the goal of counting, we would be very far in reaching our goal. But this is not the only goal of research. One of the most apparent problems is that “Big Data is not self-explanatory [...] the specific methodologies for interpreting the data are open to all sorts of philosophical debate”, as David Bollier has argued (2010, 13).

Jockers has explained that “a computational approach need not be viewed as an alternative to interpretation”⁵¹. Interpretation plays some role in all steps in a data analysis: In how we should quantify the aspects we want to investigate? And when we are provided with numbers we have to find out how the numbers translate into aspects of the objects. After quantifying we have to handle the large amounts of data to find out what to look for, where to search, how to clean, arrange and re-arrange the data. Though these processes have become much easier with technical aid, such as with visualization tools that can

50 Quoted in (Walser 2003, 16), which presents similar views as my, but nevertheless is slightly off the topic to be covered directly in the text. Walser also argues against “formalist approaches to music analysis” and objective scientific language on music, and pro the “mediation of meaning” (2003, 21).

Walser also argues “that any cultural analysis of popular music that leaves out musical sound, that doesn’t explain why people are drawn to certain sounds specifically and not others is at least fundamentally incomplete” (Walser 2003, 22). However, viewed in the context of this thesis, this statement is only a argument for why we should investigate sound. And this why is not the main focus here.

51 Within the natural sciences, human interpretation is not necessarily completely suppressed, despite emphasizing an objective ideal. As Jockers explains “[e]ven scientists will “interpret” their evidence through a lens of subjectivity” (2013, 6). Throughout the project, I have read texts from both cultures, and in my view, the main difference regarding interpretation is how much the interpretation part is accentuated and interweaved in the text. Where I, trained mostly as a humanist, have chosen to blether (would some say) about theoretical concerns, a natural scientist would probably expect to find these reflections comprised in separate discussion or perspective parts.

arrange the data instantly, each of them involve human choices: We, for instance, have to interpret and select which correlations are useful, which are unimportant, and which are even spurious⁵². And we have to understand what we have found and relate it to a context and a question. The numbers do not speak for themselves, no matter how many they are. Though we can measure multiple aspects of all elements in a corpus, search for correlations, we will still have to understand what we have found.

Therefore the computer does not render the music analyst superfluous. And when Anagnostopoulou and Buteau (2010) argue that a music analysis never can be neutral⁵³, I agree. There will always be a lot of decisions to be made. Anagnostopoulou and Buteau also state that "[m]usic analysis should not be a mere 'output' of an algorithm". This is self-evidently true, because an algorithmic output more has the character of a fact. It has to be connected to something to enter an analysis. If someone, for example, states that there are around 450 kilometers from Copenhagen to Berlin, he or she is introducing a fact. The fact can be disputed or modified. But it can also be useful for all kinds of considerations. Should I try to walk to Berlin? Should I drive there by car? Should I try to sell bananas in Berlin?

Correspondingly with audio features, they represent facts in the sense that everyone would end up with the same results if applying the exact same tool on the exact same audio file. But the measurements that are produced have to be connected to something because as Beard and Gloag explain, a music analysis "is an interaction between music itself, music theory, aesthetics and history" (2016, 14).

4.4.2 No standard recipes for neither ACA nor music analysis

The nature of ACA features further emphasizes that interpretation is required. There are plenty of ways to create ACA data, and therefore, there is no standard feature that has to be used. Consequently, there is not just one set of metadata that can be linked up with a piece of music. Rather there are all kinds of settings and choices, one has to make when applying MIR. In practice, the data most likely will be derived from the software's default

52 Vigen (2015) or the website <http://www.tylervigen.com/spurious-correlations> provides many funny examples of spurious correlations. There is, for example, high correlation between the "divorce rate in Maine" and "Per capita consumption of margarine" through the 2000's. The point is that with big datasets we can create all kinds of correlations, but we have to decide which are meaningful.

53 Anagnostopoulou and Buteau refer Jean Jacques Nattiez' (1975) idea of the neutral level of music analysis. They explain that "[t]he neutral level in music analysis is the study of a piece (or pieces) of music, without taking into account the composer's intentions (poetic level) or the listener's cognitive mechanisms, intuitions, aesthetic judgements, emotions, or reactions (aesthetic level)" (2010, 75). Nattiez claims that an analysis "can be neutral, systematic, rigorous, and scientific" (75). However, Anagnostopoulou and Buteau's main argument is that this is not possible.

features, but this can to some extent be altered according to what is desired⁵⁴. On the one hand, this large flexibility might blur an ACA informed interpretation because it makes it easier to make data say what you want it to say. And correspondingly an opponent will often be able to claim that the results could be otherwise with other settings or other metadata. But on the other hand, it also makes it easier to design your analysis, so it fits with your questions.

One could also argue that this flexibility fits well with music analysis since there is no standard way of analyzing music either. This is even truer when you analyze music with ACA methods because there is not yet established any best practices. Music analytical approaches should not be considered dogmas, and I get back up from Nicholas Cook and David Huron. Cook compares, for example, a Schenker-analysis with a large bulky tool, it “can be used in a lot of situations but not when you need a hammer” (Cook 2005, 6). While Huron declares that he sees

“methodologies as tools for conducting research, not as philosophical belief systems. Like all tools, a given methodology is suitable for certain kinds of research but not other kinds. In pursuing a research program, thoughtful scholars will take stock of the conditions of their area of research and choose a methodological approach that best caters to the research goals -- in light of the opportunities and dangers inherent in specific tasks. The most appropriate methodology may change depending on the specific task or hypothesis being addressed.” (1999, 160)

The fact, that there are many music analytical methods, or in Marsden's words “we have [...] many theoretical systems but none that is comprehensive” (2009, 141), implies that there are not particular aspects that are necessary to count. Rather it depends on the specific purpose of the analysis.

ACA methods do not simplify this matter, quite the contrary. Firstly, we simply lack experience in practicing research with them. Neither is there a perfect between the traditional music analytical methods and features. What we want to count has an opaque relation to our measurements. If we, for example, want to compare tempo in two corpora using ACA methods, we cannot be sure that ACA has measured tempo as we would.

54 The distinction between *capta* and *data*, as Drucker proposes (2011), can therefore be useful. *Capta* emphasizes the construction; “[c]apta is ,”taken” actively while data is assumed to be a “given” able to be recorded and observed” (2011, 3). In the context of ACA, one could argue that *capta* to some extent is more suitable than *data*. However, I have chosen data because it is the most commonly applied word of the two.

4.4.3 How to datafy music

Regardless whether you want to apply born-analogue or born-digital features, you have to specify what to count. Leonard Meyer offers useful ways of thinking about the relationship between music, its quantization, and musical traits. And he even does so in relation to the analysis of large amounts of music - and deploying computers for such studies:

“It would appear desirable to define as rigorously as possible what is to count as a given trait, to gather data about such traits systematically, and to collate and analyze it consistently and scrupulously - in short, to employ the highly refined methods and theories developed in the discipline of mathematical statistics and sampling theory. I should add that I have no doubt about the value of employing computers in such studies, not merely because they can save enormous amount of time but, equally important, because their use will force us to define terms and traits, classes and relationships with precision - something most of us seldom do.”
(1989, 64)

What I would like to address here is not so much the rigorous and positivist attitude, but the question of formalization of music analytical methods. The computer can ensure that there is consistency in the ways we compare pieces of music. However, we need to define at least some quantitative measures to calculate before we run the program.⁵⁵

Say, you want to compare the old with the new songs from artist Q. A common ACA approach would then be to calculate some features from audio files containing Q's songs. But since there are an unlimited number of possible features, you can extract, you have to define which to choose beforehand. Let us say, that you amongst others want to compare predominance of certain melodic intervals, the occurrences of particular rhythmical patterns, and that rhythmical timing is of interest. All these inquiries are in principle possible, and the computer can help you collect the information on all this, and it can do so very quickly. However, you have to define very precisely what aspects you investigate, and you have to define how to count them.

How would you, for example, choose to measure or count the rhythmical timing automatically? Let me think aloud: You could, for example, choose to detect onsets, the peaks in the strengths of the audio signal, and measure the lengths between them. You could choose to compare these length between onsets with the lengths between beats,

⁵⁵ These measures could be pre-defined by the software.

measured from the tempo estimation. You will then be able to calculate a distribution. Say, there is a predominance of onset lengths of 33% and 67% of lengths between beats, the song would probably then contain a lot of triplets. You could then set up a feature, which measures whether there is a prominence of 33% and 67% lengths compared to 25%, 50%, and 75%. In the former case, the song's rhythm is shuffle. You would then be able to calculate the distributions of shuffle versus non-shuffle songs in your corpus and, for example, compare the old and the new songs. And if someone else already had programmed this inquiry, it would only require you to tell the program what music to process.

But still then there are a lot of blurry factors. Firstly, how is the tempo calculated? What if the computer "believes" that the beat markers are those accentuations you perceive as 66%, the eighth notes? In that case, those onsets that are humanly perceived as 33%, would be calculated as 50% in your formula, and the song would be classified as non-shuffling. If a song alternates its rhythmical feel between shuffle and straight, it would probably be perceived as a significant musical trait, but the model would not be able to account for this. And what if the beat is divided for example 60/40, which group should it fall into then?

I am trying to exemplify and sketch out here that there are a lot of mathematical and musical factors that should be taken into account when creating just a single metric for measuring one single aspect of the groove. It takes hands-on-practice to get experience with caveats and pitfalls when understanding the measurements. And it requires you to activate both mathematical and musicological thinking. Even if someone else had developed a useful metric for this task you would not be sure if the method also is suitable for your corpus, and what the pitfalls and caveats are in relation to it.

In practice, the quantization of musical traits can become very complex and leave us with new challenges - both in the process of quantizing the inquiries and understanding the results. And these challenges of formalization are only complicated further when bringing music analysis into a digital and audio realm. In relation to music analysis' future employment of ACA methods, I see at least four questions that should be examined through practice:

- 1) The defensive question that concerns to what degree it is possible at all to define musical traits in terms of ACA features. Or put in another way; is there any meaningful and useful connection between ACA features and the music's appearance?

2) How should we, dependent on the task, consider formalizing our methods? And can we create methods that can become useful for other studies, other corpora?

3) What happens to the relationship between music, its quantization, and music analysis, when methods become increasingly complicated, numerous and customizable, which for example gives researchers more opportunities to make the data say what you want it to say?

4) To what extent can we now with statistical techniques combined with big data approaches (where multiple sorts of data can be collected for automatically finding patterns in it) apply machines for finding patterns and measure what is important? And what can we learn from such processes?

On standardized metrics

Questions 2) and 3) can lead to a discussion about standardized metrics. On the one hand, such standards would strengthen the inter-study value since it will make comparisons across studies and corpora better possible, and thereby activate even larger corpora. On the other hand, this could damage the flexibility of the methods, both the adjustment of the ACA methods, and the quantization of music analytical inquiries. In addition, the ACA methods are constantly improved, and why not apply the current best method for a task?

4.4.4 On other applications

In the swing example above, I implicitly also sought to demonstrate that the understanding of an audio feature is gradual. Either you fully understanding what an ACA feature represents, or you have some amount of intuitive understanding about it, or you have no idea about the connection from sound to feature. Though it will probably always be impossible to completely understand the precise translation from the audio file to ACA feature, unless it is trivial⁵⁶, more knowledge about the applied feature is always preferable. But sometimes no knowledge may suffice. It depends on the purpose. Even if this obstacle of formalization becomes too hard to handle, it is not necessarily enough to argue against deploying the methods.

As I explained in Chapter 3, there are a lot of other potential purposes of applying statistical techniques: They can be applied for exploring a corpus, describing music or quantitatively confirming theories (Nettheim 1997, 2). The question of formalization is

⁵⁶ An example of a trivial feature could be whether there is sound in the signal or not.

most important for the two latter cases. When you are creating statistics for a descriptive or confirmatory analysis, you will want to create quite precise inquiries in order to understand what is measured in the music. However, detailed understanding of formalization is not necessarily required if you deploy ACA methods for exploring a corpus.

Statistical methods can, for example, help us arrange our data, sort it, map the music, cluster it, help us to make rough identifications. Thus statistics can be used as a means to reduce complexities, as "*a potentially useful strategy for discovery* rather than a *belief about how the world is*", to apply the words of Huron (1999, 132, italics in original). Or as Tukey remarks about exploratory data analysis: "Anything that makes a simpler description possible makes the description more easily handleable" (Tukey 1977, preface).

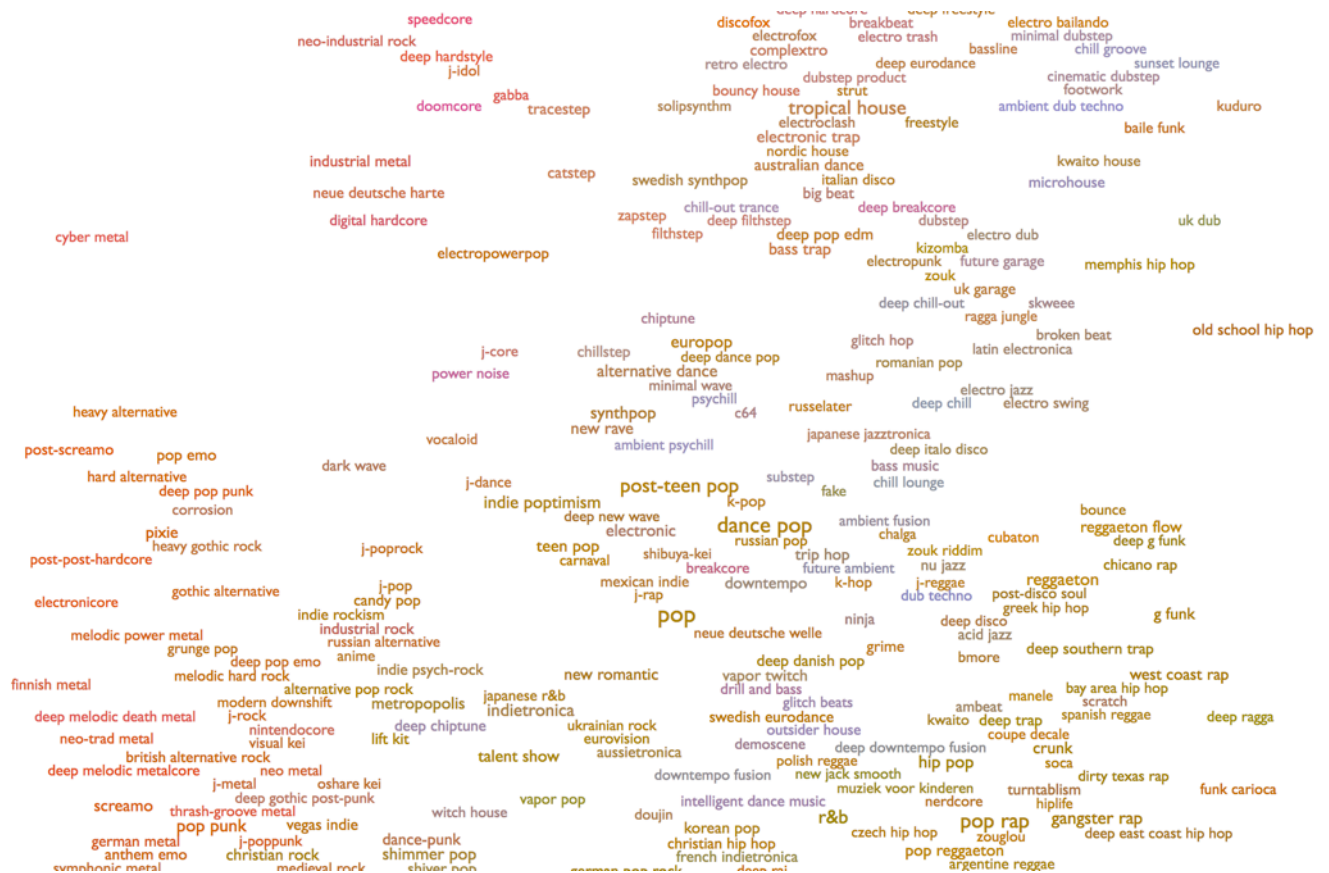
The website Every Noise at Once⁵⁷ (Figure 6) is a good example of a rough mapping of a vast amount of music, classified as music genres. It is to my best of knowledge mapped on the basis of Echo Nest's automatically extracted audio features. At this moment of writing⁵⁸, the website shows 1533 music genres mapped according to their musical characteristics. Each genre contains a "sub-map" of artists adhering to the genre. You can also search for an artist in order to find out which genres he or she is ascribed. Even though there apparently is little consistency in how genres are labeled or created⁵⁹, and how artists⁶⁰ are assigned to them, this map could be a useful tool for exploration. If you are curious about a genre, the website can become a tool that provides you with an introductory overview of artists within the genre. The detail level at Every Noise at Once is very coarse-grained, and the website is too inflexible to be able to deliver the musicological details most likely required. But the map demonstrates that ACA methods can measure in some accordance with how music sounds. It displays that the answer to question 1) above in Section 4.4.3 is that it *is* to some extent possible to define musical traits in terms of ACA features. Correspondingly, it displays that ACA features can provide some measure of music similarity, which is not arbitrary.

57 www.everynoise.com, retrieved December 15, 2016

58 June 13, 2017

59 "modern classical", "modern performance", "violin", "baroque", "deep classical piano", "liturgical" are examples of genres in the classical area of the map. They demonstrate well the lack of consistency in genre labels. Some of these genres represents periods, some represent instruments and some represents contexts. <http://everynoise.com/EverynoiseIntro.pdf> provides a description of the creation of genres, retrieved December 15, 2016.

60 In the case of classical music, "artists" comprise of a blend of performers, orchestras and composers.



4.4.5 How to incorporate ACA in musicological practices

The *Every Noise at Once*-map exemplifies well how ACA can become a tool for other inquiries, a jumping-off point to assist discovery, a tool that can suggest us where to listen. But computers do not necessarily entail this limited amount of flexibility for users. They can also enable us to modify inquiries, look at our data, listen, re-formalize measures. They allow us to do so in some recursive process, where one step can inform the next. Thus these types of processes can also “help to refine listening experience”, for example may visualizations “instigate [...] a more informed and searching experience of the music”, as Cook advocates (2010, 12).

Jockers suggests that we apply computers with a “blended approach”, blending between close and distant reading, in which “the two scales of analysis work in tandem and inform each other” (2013, 26). Machine readings on the macro level can create better understandings on the individual work level. Ørmen and Thorhauge (2015) have provided a useful example of the opposite direction creating mutual benefit; thorough analysis on

the micro-level can create a better understanding of what is actually counted, and thus inform the macro-level analysis. Thus, digital methods can help us zoom in and out of levels, from overviewing the many to scrutinizing the individual and back again. In this view, macroanalysis is not a replacement for traditional close reading. Rather a supplement. Whether to apply quantitative macroanalysis or close listening depends on the question and purpose.

However, at the current state of ACA methods musicology, we must improve the understanding of what the results on the microanalytic level entail. We lack experience interpreting the data. It is not sufficient to rely solely on quantitative data. Rather, it will, to understand the results on the macro-level, be necessary to couple them to knowledge on both qualitative and quantitative levels on the micro-level.

With Figure 7, I suggest that there are (at least) three levels that should inform each other reciprocally. In the bottom right corner is depicted the traditional close listening mode of manually listening, which implies engaging of music analytical listening skills. The bottom double-arrow serves to demonstrate that you will have to understand how the acoustic qualities convert into quantities, and oppositely also what aspects of the music that correspond with the features. This arrow corresponds with the approach I in Section 3.4 labeled *methodology-informed* listening. The left double-arrow involves sheer quantitative reasoning. After we have measured the individual audio files, the next step in the large-scale analysis is to calculate some statistics of the many files. It will often be desirable to visualize these results to grasp them better. The right double-arrow demonstrates that the results on this macro-level have to be understood in relation to how the objects were quantified in the first place. It will most likely be necessary to apply domain knowledge in this process.

This triangulation also constitutes my approach to the methods throughout my case studies. The essential question is how to translate between these different ways and modes of grasping the objects of analysis. Especially throughout my analysis in Chapter 7, I will pursue making this process, of what Kirschenbaum labels “rapid shuttling” legible (2009). This approach entails alternating “between quantitative information and traditional hermeneutic close reading” (Hayles 2012, 47), and I will pursue elucidating it by demonstrating both my and the machine’s steps in the analytical process.

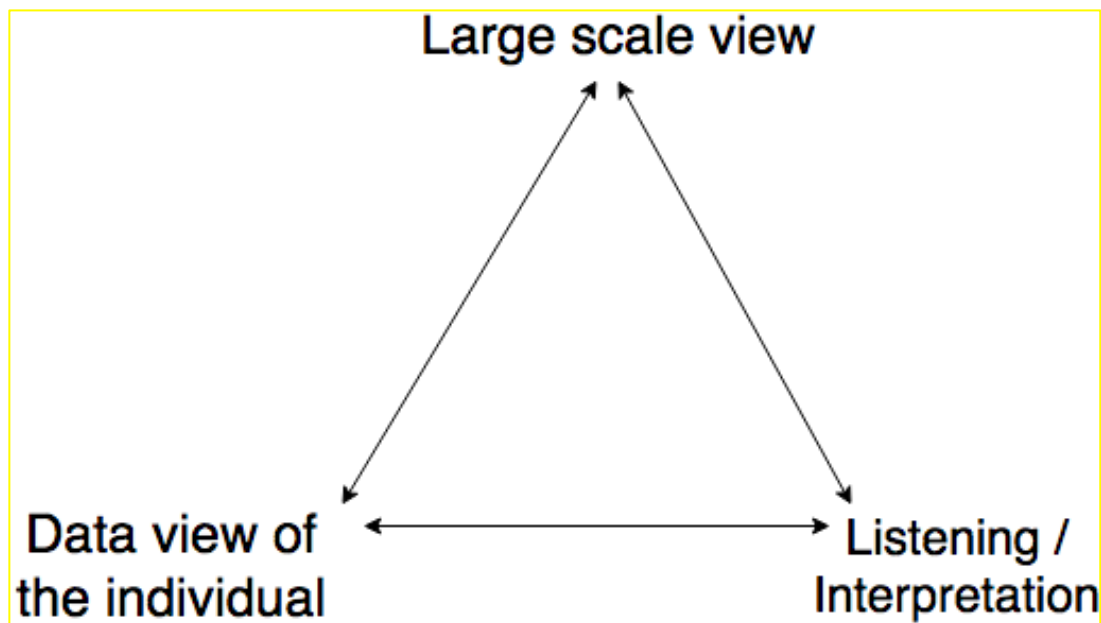


Figure 7 Three different “views” on the objects. These views ought to inform each other.

4.5 Can We Trust Data?

The model presented above is not only useful when practicing, it is also useful when thinking about the epistemological value of the techniques. Firstly, it can be helpful to think about what it does to the object of study that it has to be viewed through “the lens of data” (Aiden and Michel 2013). Secondly, there are specific issues tied to ACA and music analysis that arise. Concerning big data analysis, Dalton & Thatcher address the matter that “we must ask what it means to be quantified in such a manner, what possible experiences have been opened and which have been closed off?” (2014, #2). This question is also pertinent to address in a music analytic context.

Especially questions about translatability arise in continuation: How good is the translation between the certain data-view applied and the analyzed object? How much of the music do we capture in the process of quantification? What theoretical considerations should we do before applying ACA? To what degree are ACA techniques capable of investigating both the surface and deep level? And how good are the views from the surface macro perspective and the deep micro perspective? Etc. I will consider these questions in my case studies. But for now, I will theoretically reflect on some general concerns in the process of translation: How to convert from music to data (Section 4.5),

how to go from data to analysis (Section 4.6), and on the epistemological value of data-driven approaches (Section 4.7).

4.5.1 Data is a partial representation of the object

To elaborate on these questions, I will take one step back and elaborate on data. One characteristic of data is that “data is always a partial representation of the object of study” as Christoph Schöch has remarked (2013). Data represents a reduced way of looking at the object. When you measure swing as exemplified above in Section 4.4.3, you are quantizing a very particular aspect of the music. Doing this, you are only looking at some rhythmical aspects, and not including other significant aspects such as accentuation or the positioning of the onsets. You are sheerly identifying where the onsets are relative to the tempo and calculating some statistics. However, this is not a new situation to musicology. Scores, likewise, represent a reduced view of the music because they primarily embody tonal and rhythmical content (Cook 2005). Audio features reduce music in other ways than scores. Schöch argues in continuation that

“this “disadvantage” of partial representation is small compared to the fact that digital data can be transformed, analyzed, and acted upon computationally.” (2013)

4.5.2 Data complicates

At the same time “[d]ata add complexity to the relation between researchers and their objects of study” (Schöch 2013). When we calculate features from a piece of music, the audio file has to be processed a number of times. Each of these adds complexity to the analysis process:

First of all, the music necessarily has to be represented digitally as an audio file. The full control of the analysis is already lost at this level since it is hard to grasp exactly how the translation between sound to audio file occurs. The audio format is one issue here. It implies that the digital representation of a song is different dependent on its audio format (Marsden 2016, 7). An audio file compressed in a lossy format such as MP3 will produce other ACA features than a WAV file. Thus the audio format influences the feature calculation. Many musicologists will not be used to dealing with these aspects of audio format, and how they influence the music analytic aspects.

The calculation of features is yet another step that complicates the transparency. If the feature is low-level we do not have prior experience in interpreting them, they have a low explanatory level, and they will perhaps not help us answer the questions we have regarding the music. Moreover, they are dependent on aspects of the mastering, such as equalizing, compression and loudness, which often will be irrelevant for the analytic purpose. On the other hand, if the feature is higher-level additional and even more complex issues arise because we will often not be sure that the algorithm fits all cases. For example, does tempo estimation not work 100% in accordance with how most humans perceive the tempo. If you apply an ACA-based tempo metric for your large-scale analysis, you append another element of uncertainty because you cannot be sure of what the features actually represent.

4.5.3 What is measured?

Another aspect that blurs the data view is that the methods are not entirely reliable. They do not measure what they seek to measure. Mainly because of music's complexity "[t]here seems to be a *glass ceiling* for many MIR tasks that lies (for audio tasks) at around 70% accuracy" (W. Bas de Haas and Wiering 2010, 177—italics in original). In addition, the features relate to the music in opaque ways, as I have argued for above. This makes it difficult for the music analyst to find out why and where discrepancies occur.

It is not only humans who do not agree with machines' output. Even music experts do not agree with each other, even when it comes to comprehending simple music analytic aspects. For example, de Clercq and Temperley (2011), whom I consider music experts, transcribed chords of 100 songs by hand. However, they did only agree on about 90%-95% of the chords depending on the task⁶¹. This complicates the idea of a music analytical "final truth." From one point of view, digital techniques only blur this dubiousness even further. When machines only imprecisely simulate something that humans do not even agree upon, it will only make the results even more imprecise compared to human perception. If there is no final truth, how can a machine even calculate it? On the other hand, one could argue that if the machine calculates it, it must be objective in some sense, since the results will be reconstructable the next time it tries. The results will this way be comparable with other similar (mis)calculations.

⁶¹ See (de Clercq and Temperley 2011, 59) for the level of agreement on different tasks.

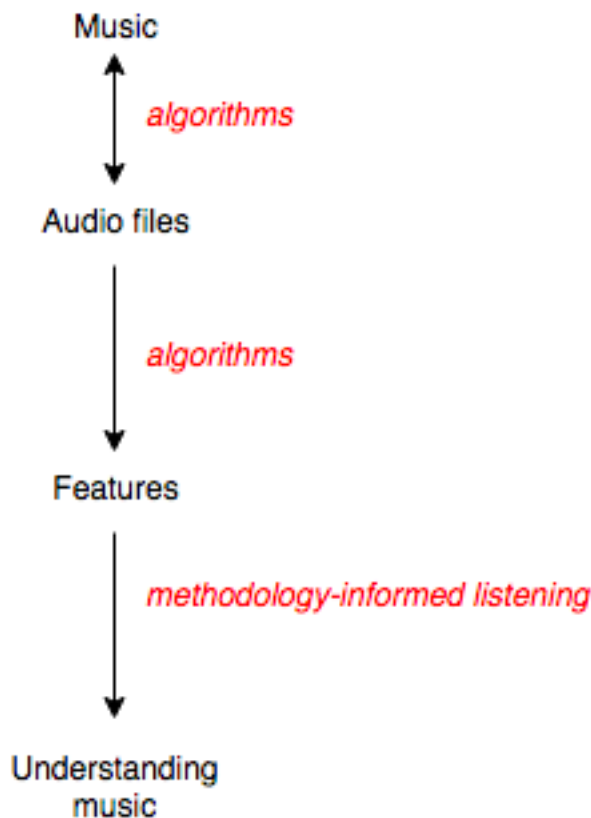


Figure 8 A music analytic diagram of the flow, from music through the creation of features to understanding the music. Each step adds complexity to the object of study. In this thesis, I will apply *methodology-informed listening* (see Section 3.4) for interpreting the features.

4.6 Can We Trust the Analysis?

4.6.1. Can we trust large-scale analysis?

But where do these indeterminacies leave the large-scale analysis? It depends on the purpose to what degree you can live with error. Notwithstanding, all these elements of insecurity and opaqueness do not render the methods useless per se. As Marsden explains, other fields, such as the natural sciences, are used to dealing with imprecise data (2016, 9–10). Therefore one part of the solution to the problem is that “computational music analysts [should] recognize that they operate with approximations” (2016, 9). This can be sufficient for viewing our objects from a distance.

When conducting large-scale analysis, we will usually want to view our objects from a distance, and approximations will often be sufficient. As Mayer-Schönberger & Cukier explain we can often live with messy data as long as we have enough of it:

“Often, big data is messy, varies in quality, and is distributed among countless servers around the world. With big data, we’ll often be satisfied with a sense of general direction rather than knowing a phenomenon down to the inch, the penny, the atom. We don’t give up on exactitude entirely; we only give up our devotion to it. What we lose in accuracy in the micro level we gain in insight at the macro level” (2013, 13).

Thus the errors cancel each other out on a large-scale, and this is one of the big advantages having large amounts data.

However, there is a small “but” incidentally mentioned in a parenthesis on page 34 that goes “(as long as they do not introduce a systematic bias)”. Systematic biases are one of the major pivotal points in big data criticism, and rightly so, because they may end up distorting these “senses of directions”. In the case of ACA, most researchers have not yet music analytical experience applying the techniques, and they should be very alert for potential biases. First of all, researchers interpreting ACA results should consider the methods applied, and what aspects they actually measure in the music. Preferably, they should also empirically investigate them, through representative sampling, for sheer theoretical considerations are not able to include all aspects of how complicated algorithms translate into measurements. Questions such as, how does mastering affect the results? should be a standard enquiry when understanding ACA features. A louder mixed hi-hat may influence the machine’s results profoundly, and in an insidious manner.

Sometimes the calculations are obviously wrong compared to humans’ perception of the music. In these cases, you should still consider and investigate the errors and reflect upon how this affects large-scale results. Again, either you might be able to live with indeterminacies because the errors cancel out each other on the larger scale, or there will occur systematic biases. A wrong tempo calculation may, for example, often be estimated by the machine as either half or double the “correct” tempo. If the particular tempo estimator performs less well on certain types of music, it may stir up the results of e.g. a tempo curve through the times. During my case studies, I have found a general tendency that ACA methods have difficulties calculating tempo in syncopated rhythms. Furthermore, the tempo of hip hop is often estimated as double the tempo, I would

perceive it. If this is a systematic trend in the algorithms' estimation it would imply that the larger the share of hip hop the larger the error.

If you have the opportunity of choosing the algorithm for your task, a dilemma, therefore, will arise. If you apply the newest and most efficient algorithm, you may, on the one hand, reduce your error, while on the other hand, you may choose an algorithm that has inherent systematic biases towards certain kinds of music. The problem is that these eventual biases have not been identified yet. Therefore they will need your attention and this will speed down the process.

4.6.2 Epistemological value

However, calculating a lot of ACA features of a corpus allows one quickly to browse through the data. Visualizing the data can be a great assistance for this, as it can help overview the objects through the lens of data. This easiness of inquiry that visualization software entails is one of the greatest practical advantages data allows. It enables you quickly (and instantly if you have arranged your dataset properly) to go back and forth between the three views presented in the triangle, Figure 7: From looking at data, over-viewing the many, listening to the music, and examining how the metrics translate to musical aspects. You can examine subsets of the corpus, and visualize trends through graphs and diagrams. In short, data can assist and activate the music analytical thinking.

In many cases, two opinions are better than one and machine can offer a very qualified second opinion. You can ensure a better alliance between objective and subjective results. However, there is also an apparent risk of ascribing too much epistemological value to the results, just because they are produced computationally. Rieder and Röhle have identified five challenges of digital methods, among which one of them is labeled “the lure of objectivity”. They explain that

“the interest in computational tools might [...] indicate a desire to produce knowledge that can compete with the natural sciences on their own terms, by being as ‘objective’, as ‘rigorous’, with the help of machines” (2012, 73)

However, they continue by questioning this ideal:

“While the ‘plodding’ capacities of machines can be usefully integrated into many kinds of research, they should not be taken to guarantee a higher epistemological

status of the results. [...] [O]n an epistemological level, [machines] create complications rather than resolving them. Questions of bias and subjectivity, which the computer was thought to do away with, enter anew on a less tangible plane – via specific modes of formalisation, the choice of algorithmic procedures, and means of presenting results.” (2012, 73)

When analyzing large amounts of music, further epistemological complications arise because it becomes easier to make data say what you want it to say, more or less consciously. Firstly, because it becomes possible to create vast amounts of metrics, and you thereby can increase your chance of finding correlations, and consequently also the risk of finding spurious correlations (Meyer 1989, 58). Secondly, the flexibility of both ACA methods and the plethora of music analytic methods combined with the ease of inquiry makes it easier to adjust thresholds or choose another music analytical metric that makes the data fit with the presupposed theory. Consequently, there is a risk that the subjectivity just changes medium and appearance; interpretation takes the disguise of being objective.

The solution to this dilemma is, as I see it, not to reject using ACA, but rather to apply them critically, aware of as many concerns as possible. This awareness could be incorporated in how you present and include ACA results in the research. For example, as David Huron advocates, by distinguishing whether you have applied the methods in the “context of discovery” or in the “context of justification” (2013, 5). Because

“[l]arge data sets require a careful delineation between a priori and post hoc theorizing. [...] With post hoc theories, one cannot legitimately use the language of prediction that is the essence of hypothesis that is the essence of hypothesis testing.” (2013, 5 & 6)

4.7 Can We Trust Data-Driven Approaches?

Not only humans can search for and find patterns in datasets, computers can also be applied for finding them automatically. For example by using machine learning methods, which is a standard technique within MIR. Say, you have a dataset of audio files with annotated genres, you can have the computer calculate how different audio features correlate with each genre. How useful are a given timbre features for predicting genre for example? How well are the tonal features calculated? Etc. The machine can hereafter

"transfer" how features correlate with genres to an even larger corpus, and thereby estimate genre of other audio files not yet annotated.

I have chosen to cover machine learning techniques here for two reasons: Firstly, because the genre example also can be conducted with all sorts of other annotations. Also continuous measures. It is, for example, possible to create born-digital high-level metrics that can estimate how we will perceive a piece of music. You can create a measure of *relaxation* by asking a lot of people to rate on a scale from 1-10 how relaxing they find different pieces of music. The algorithms can then calculate how and how much different audio features correlate with the ratings, and after that estimate the expected relaxation of other pieces of music that have not been rated. Viewed in the light of Mendel's categorizations of music analytic questions we can pose with computational techniques (see section 4.2.4) this method is a way to more directly approaching the aesthetic level⁶²: It is a way to calculate how aesthetic properties of the music comprise of and correlate with properties of the sound.

4.7.1 New views on music

Secondly, machine-learned features are also interesting because they form a way to re-examine how we as humans grasp music. They can teach us something on what roles different musical aspects play in how human perceive music, how we structure it, etc. MIR focuses mainly on the model building process and on reporting how well the model performs (J. A. Burgoyne, Fujinaga, and Downie 2015). However, from an analytical point of view the intermediate results can be interesting; what factors and weightings constitute the individual results. If provided with a dataset of Kenyan and Tanzanian recordings, we can ask the computer to try to learn which audio files go into which category from analyzing the audio features. Next, we can investigate which features the computer associates with each country, and thereby learn about how a computer interprets the difference in music between the two countries. This enables us to get insights in which features, and consequently also which musical aspects that are most identifying the music from the respective countries within the corpus.

Nicholas Cook has explained that most music analytical methods, although they seem very different, ask the same questions, "how components of the music relate to each other, and which relationships are more important than others" (1987, 2). Machine-learned techniques are heuristic, but they can be an approach to investigate the latter question of

62 The aesthetic level regards modes of expression

importance. They comprise a way of trying to more objectively measure what counts musically according to a particular setting and task. Therefore they can potentially offer us new perspectives and insights in which aspects matters in music. For example what musical aspects are characteristic of a certain genre, a dance hit, a particular composer's style, in relaxing music, or many others of this type of inquiries.

This way of thinking and posing questions through data-driven approaches data is exemplified by Hallinan and Striphas (2016). They have analyzed the implications of the Netflix Prize, which was a competition intended to boost the efficiency of Netflix's movie recommendation system. Though setup as a technical challenge Hallinan and Striphas amongst other argue that the point of departure, a large dataset containing mostly of user ratings, in combination with the competition task, to create the best recommendation algorithm, led to suggestions of "new models of cultural identity latent in the dataset" (123). What is especially interesting is that the new models "tended to reject dominant [...] demographic categories in favor of emergent frameworks of identification" (123) as if "the machine may be understanding something about us that we do not understand ourselves" (123).

It is hardly any surprise that demographics such as *age*, or *gender* only can account for crude tendencies in our consumption of culture, and that reality is much more fine grained and complex. Nevertheless, Hallinan and Striphas still show that data-driven approaches can offer us new ways of viewing cultural objects and how we conceive them. Alan Marsden (2009, 145–46) makes a similarly point analyzing the confusion matrix⁶³ of systems for automatic genre classification (J. Downie et al. 2005). He explains that though "[t]he objectives in this research were not explicitly analytical, [...] analytical conclusions can be drawn" (2009, 146). Marsden analytic conclusions are, however, somehow vague and suggestive: "punk music appears to have the most distinctive sonic characteristics; 'new age' music has some characteristics which are similar to 'ethnic' music, causing some systems to mis-classify, but other characteristics which are distinctive."

4.7.2 Analytical limitations of datadriven models

Marsden applies word *appears* to describe the relationship between the correlation matrix and the music analytical statements. I agree in using it. Economists also tend to use the word *indicate* to accompany their data analyses, because data stands in for the objects of

⁶³ A genre confusion matrix shows the initially labeled genre and how many times the algorithms predicted it rightly, and to be any oth the other genres. If, for example, *jazz* and *punk* music were two of the genres, you will be able to see in how many instances songs labelled *jazz*, were predicted as *punk* music by the algorithms. And vice versa.

study. More generally, the analytical implications that can be drawn from data driven approaches depend on 1) the parameters built into the model, and 2) how the model has been trained. Therefore we should be careful of concluding too much from these techniques.

1) A model is dependent on the data put into it. However, no amount of data is currently able to include all relevant musical aspects. But since the data cannot capture the music in its entirety, the model will not be able to include all sounding aspect that counts:

One aspect of this problem is that many events happen simultaneously, and any of the events happening at a particular time may be relevant for the perception of the music. The consequence is that for the model to capture everything happening, it has to be very high-dimensional. De Haas and Wiering explain:

“Music is a complex phenomenon; therefore a considerable number of musical features need to be taken into account at any given point in time. For example, a MIR system may need information about simultaneously sounding notes, their timbre, intonation, intensity, harmonic function, and so on. As a result, the input vector, i.e. the list of numerical values representing these features, is often high-dimensional“ (2010, 178).

Music’s temporality complicates the number of calculations to an incomprehensible extent. In music, the position and accentuation of events matter a lot. And they matter in relation to a multitude of levels simultaneously; the overall elapsed time, their relation to adjacent musical events, what has happened before in the piece, expectations, the specific sonorities playing the events, etc. It seems impossible to calculate all significant factors into a model. Also, a method that seems able to account the most important aspects of one song might not account for another song’s most important aspects.

In practice, when interpreting ACA results, unless the methods built into the model are explicated, it will probably often be very simplified models of music change that are build into the model. If time is taken into account all, which it rarely is, as de Haas & Wiering also explain:

“For example, when dealing with audio data, a common paradigm is to split an audio file up into small (overlapping) windows. Subsequently, a feature vector is created for each window, which contains characteristics of the signal [...] These feature vectors are inputted into a classifier for training and the temporal order of

the feature vectors, and thus the notion of musical time, is lost in the process. In a sense this resembles analyzing the story in a movie while randomly mixing all the individual frames” (178).

A third obstacle when attempting to capture everything in the music into a model relates to both simultaneity and temporality. De Haas and Wiering refer to it as “Not All Information Is in the Data”:

“[M]usic only becomes music in the mind of the listener. Hence, only part of the information needed for sound judgment about music can be found in the musical data. An important piece of information that is lacking in the data is the information about which bits of data are relevant to the musical (search) question and which bits are not, because this is often not clear from the statistical distributions in the data. For instance, in a chord sequence not every chord is equally important (for example passing chords or secondary dominants) and a harmonic analysis of the piece is needed to identify the important chords. Similarly, most musically salient events occur at strong metrical positions, and a model is needed to determine where these positions are” (179).

De Haas & Wiering hereby addresses, as Bob Sturm expressed it in a blog post, that “[t]he sampled audio signal is only half of half of half of the story” (2012). Though Sturm addresses MIR researchers, the quote also elucidates ontological limits of ACA analysis. There are a lot of processes, such as cultural or psychological, that the sole information in the audio signal cannot account for. Though these limits principally are out of the scope of this dissertation, they have consequences that reach into the ACA analytic realm. Our analysis (be it manual or computational) of a piece of music is affected by how we perceive the music, and certain chords, for example, will have larger effect than others. Just as small changes in accentuation can alter the perceived groove.

These obstacles of ACA model building imply that if a model, for example, performs well on timbre features and less well on the rhythmical features we can conclude that timbre characteristics play a role in relation to the given task. But not that rhythm necessarily is insignificant. The results only indicate that *this model's* particular way of handling the timbral features was better than the model's particular way of handling the rhythmical features on this specific dataset. So, the direct analytical value of analyzing the intermediate step is rather questionable, and it is hard to reach final conclusions. And that was probably also why Marsden used the word appear.

2) The dataset and its labels used for classification constitute a second constraint regarding what we can learn from data-driven approaches. At the end of the day, the results are only as good as their datasets; they are restricted by the quality of the so-called ground truth, how well the dataset corresponds with its labels and the representativity of it. This is a vulnerable point in data-driven approaches regarding music. The method's inherent weakness is that it is "actually just exploiting confounded characteristics in a test dataset" as Sturm and Collins have put it very precisely (2014, 1). If all songs in a dataset that are labeled rock have a tempo of 122 BPM, the machine will perhaps "believe" that all other songs with a tempo of 122 BPM also should be ascribed the label rock. This example is extreme and simplified and concerns a dataset of very questionable representativity, but nevertheless, this kind of thinking should be invoked in any case: Datasets are always biased in certain directions and so are the annotations, and these biases affect the machine-learned results. One of the current typical biases is that datasets contain western music and western annotations. They thereby represent westerners' views on music.

When conducting machine learning tasks, it is crucial to have a good and representative dataset. These are not easy to get though because the creation of music expert datasets is a "costly and time-consuming enterprise" (W. Bas de Haas and Wiering 2010, 178).

However, some have been created and made accessible (e.g. J. A. Burgoyne, Wild, and Fujinaga 2011; Bertin-Mahieux et al. 2011). However, especially the vast amounts of information created on internet has enabled many new enquiries concerning large amounts of music. We can get a lot of useful information from users' tags⁶⁴, datasets with genre information⁶⁵, music blogs⁶⁶, users' comments⁶⁷, reviews⁶⁸, and databases with lyrics⁶⁹, etc. These sources allow a myriad of new queries such as, are there correlation between usages of certain words in the lyrics and musical aspects, and how has this developed through the times?⁷⁰

Another negative side effect of the restrictions of ground truth data is what de Haas & Wiering refer to as Danger of Overfitting. This has to do with the flexibility of the model:

64 e.g. www.last.fm

65 e.g. www.allmusic.com

66 www.hypem.com hold a collection of music blogs

67 e.g. on www.youtube.com

68 e.g. www.albumoftheyear.org/

69 e.g. www.musixmatch.com

70 For the technical novice, machine learning tasks may sound very challenging. And it is. However, it has also become easier for non-programmers to perform such operations, for example by using software tools such as Weka, "a collection of machine learning algorithms for data mining tasks." <http://www.cs.waikato.ac.nz/ml/weka/> - retrieved February 22, 2016

“Obviously, the more flexible a model is the better it can fit the data. As a consequence, a flexible model will often have a larger prediction error on other data sets than a less flexible model because it was trained to explain the noise in the training set as well” (2010, 178).

One problem arises when the flexible model is applied for calculating classifications on other datasets. A very flexible model that fits very well with the ground truth data, may be less precise on other corpora, because it is trained to account for the differences in the particular ground truth dataset. Consequently, as de Haas & Wiering also explain, “it is often unclear if these systems present an improvement that can be generalised to other data sets, or if they are merely overfitting the currently available data sets” (178).

The implication of this is that a model that works best and is evaluated on one kind of corpus not necessarily is the model which is best suited to the style of music intended for analysis. A model that has high accuracy in estimating the tempo in a corpus of rock music does not necessarily perform well on Indonesian folk music or western classical music.

4.7.3 How far can we take analytical conclusions based on data-driven approaches?

Consequently, it would in many cases be more appropriate to formulate more carefully and defensively what data-driven methods can teach us. These methods at the end of the day only verify that some musical aspects actually counts in a very specific dataset. And therefore there is a long way before we can state that we can precisely measure the strength of the artist or genre signal, for example.⁷¹ One of the things that Ann Ova and Mel Cepstrum, in their fictive dialogue, find that MIR and music psychology have in common sums well up a cautious approach to these types of analyses (Aucouturier and Bigand 2012). The two researchers end up agreeing that analyses of the type “proof of feasibility” is a field of shared interest. For example, if a machine can show “that there exist enough harmonic information in [...] Indian classical music [...] to explain the good performance of western listeners when they are asked to classify emotions in raags” (401)⁷², we can prove that the information exist in the signal. Accordingly, one can imagine many inquiries of this type; for example, whether a machine can discern between female and male songwriters, independently of the singer’s gender.

71 Jockers tries to measure these factors, and I find the results and his approach very interesting. But I would also like to see them calculated with other methods. (Jockers 2013, chapter 6)

72 Referencing (Balkwill and Thompson 1999)

On the other hand, dismissing these methods sheerly due to their constraints would be to neglect that approximation and reduction is a useful strategy for discovery and knowing. The data-driven process is similar to what music analysts always have done; searched for and accounted for aspects in the music that appear meaningful. While never being able to reach final conclusions about what aspects precisely are the most important. George Box' famous quote saying:

“Essentially, all models are wrong, but some are useful” (Box and Draper 1987)

is indeed valuable both for manual and data-driven practices. A map is a good reminder of this: It is essentially wrong, but very useful for many purposes.

4.8 Rounding off Chapter 4

In this chapter, I approached my research question theoretically. I covered those issues I find it most relevant for my target group to know: I delineated what ACA methods can calculate, and how they do it. I outlined some of the prospects they enable, arguing that they can enhance our perception, enable new modes of investigating, create better bases for theories, etc. And I discussed how they could be incorporated in humanities' practices accounting for particular music related concerns.

The theoretical notions, I presented, were overriding, and deliberately so. Their purpose was to set up a theoretic framework that can serve to sharpen the attention when deploying ACA techniques. These points of attention will also guide how I approach and discuss the empirical studies: I consider the advantages of conducting large-scale analysis presented in 4.2 as goals or aims which I can discuss the specific methods potentials of fulfilling. The notions presented in 4.3 and 4.4 on musicologists can incorporate the tools into their practices will be integrated into how I approach the tools and evaluate their value. In the last part of this chapter (4.6 - 4.8), I asked whether we can trust the methods. I did not provide any very clear answers, but rather I lined out overriding points of attention that are more generally applicable. The next chapters will serve to provide more concrete answers to these questions.

CHAPTER 5

Echo Nest's Features

Bridging from Machine Learning to Musicology

The next three chapters will contain empirical case studies, in which I will investigate the methods and their applicability for musicological purposes. These case studies span from exposing primarily methodological concerns towards applying ACA methods for a music analytic case. I will implement the theoretic knowledge, which I have presented in the previous chapters to guide me where to focus my attention towards the questions that are most pertinent to address in relation to the target groups' needs. Chapter 5 will be a mainly methodological investigation of Echo Nest's high-level machine-learned metrics. Chapter 6 will discuss the epistemological value of an existing large-scale music study. And in Chapter 7, I will investigate the features by practicing music analysis with them.

(This chapter is a rewriting of (Andersen 2014). All examples of songs that I mention are gathered on a Spotify playlist: <https://open.spotify.com/user/bvinderusland/playlist/0uwDlprNHLMVZ816olkwTE>)

5.1 Introduction to Machine-learned Features

Burgoyne et al. have asserted that lower-level⁷³ features calculated very directly from the audio files are “necessary to process audio but [...] not especially interesting musicologically” (2015, 222). I suppose they are referring to the difficulties of translating from very direct measurements of the audio files to musical characteristics or qualities. Burgoyne et al. suggest instead that “high-level-tasks would be of greater musicological and cultural interest” (222). Thereby they insinuate that it can be desirable to perform

⁷³ See 1.7.3 for an explanation of low-level and high-level features

some kind of mathematical operations on low-level features to create features that correspond better with human perception of music.

The apparent challenge with music analytic inventories at low levels is that at the lowest-level of feature generation is that we are simply unsure how some measurement correspond with our perception of the music and whether it is at all relevant for our understanding of the music? What does it, for example, mean if the Spectral Flux⁷⁴ is high in a song? Can it enhance our understanding of the music? And to what degree is this measure at all music analytically useful? Another challenge is that the lowest level features calculate aspects of the music on such fundamental plans that it takes some mathematical modeling combined with psycho-acoustical knowledge to calculate even the most basic of the traditional and well-known musicological concepts from low-level features (see e.g. M. Müller et al. 2011).

Even if we have calculated commonly applied measures in music analysis, it can be hard to translate them into qualities (as argued in Clarke and Cook 2004, 6). A concrete example of this is found in continuation of Schellenberg and von Scheve's study (2012).

Schellenberg and von Scheve manually measured the tempo and the mode in several songs that have occurred on the Billboard Chart's top 40. In the conclusions, they suggest that there has been an increase in mixed emotional cues. This reasoning is based on their measurements, which showed that among the fast songs the share written in minor⁷⁵ mode has increased during the last decades. Correspondingly, a smaller share of the slower songs is now written in minor than 50 years ago.

Schellenberg & von Scheve themselves note that tempo and tonality are perhaps the most important measure of happiness or sadness. I do not want to dispute this premise, nor their approach. Rather I will point towards the fact that this is a point where computational methods can provide us an extra view: Computers would not only have allowed Schellenberg & von Scheve to examine a larger sample set. They would also have enabled them to include more parameters to confirm or nuance the findings. Does it, for example, affect their conclusions if popular music has become increasingly rhythmically focused, as I would assert it has become?

ACA methods would also have allowed an approach to answering the question through the modeling of emotions. Data-driven approaches (Section 4.7) can be applied to investigate correlations between the annotations and the audio files' feature: If we have a dataset of

⁷⁴ Spectral distance between successive frames (Lartillot 2014, 60).

⁷⁵ The mode was "defined as the mode of the tonic triad" (199).

audio files with annotated happiness-ratings from 1-10, one approach could be to determine which features correlates most with happiness and sadness. Next, it would be possible to investigate the development in these features. This method would be able to deal with the fact that tempo and tonality are not the only indicators of happiness. By taking more factors into account, the approach would be better able to account for the complex relation between music and emotions. Yet another method would be to start with the initial dataset with happiness-ratings, and after that apply machine learning to estimate the happiness of all the songs in the corpus. If these happiness values were becoming less spread-out and more centered (i.e. if the standard deviation has decreased) it would be a further indicator that Schellenberg & von Scheve have a valid argument that there has been an increase in mixed emotions in hit music.

I chose the example to demonstrate that there is a complex relation between qualities and quantities. And though this gap between in practice never can be closed, data-driven approaches can become one way of diminishing it and handling the complexities. Computers can assist us simulating music's qualities by combining quantities that correlate with qualities. This implies that we can use the methods to try to calculate any kind of musical aspects. We can create our own metrics with any qualities we can imagine, such as "relaxness," "aggressiveness," "rockness," "intro-quality," "Scandinavian-ness," "rhythmical complexity," etc. However, these metrics will always be limited by our ability to model the qualities, which in most cases will be constrained by the ground truth, the initial annotated dataset with its features, as argued in Section 4.7.

5.2. On Echo Nest and My Purpose

5.2.1 Echo Nest

This approach has proven useful to Echo Nest (EN), a self-proclaimed music intelligence company, which calculates audio features for various tasks⁷⁶, such as playlist recommendations (Whitman 2013). Echo Nest have created a series of features they found useful for their purposes, and they have gradually refined these features to improve them. At the time of my investigation, The Echo Nest API contained automatically analyzed data of more than 36 million songs within a very wide variety of music genres. It contained

⁷⁶ In their own words: "We help music companies develop and commercialize the most advanced, personalized and engaging music applications in the world." <http://the.echonest.com/solutions/>, retrieved January 24, 2017

more than 1,2 trillion data points in total⁷⁷. These data points include both automatically derived values of acoustical features and song metadata, such as the name of the artist, album, among others. It is currently possible to engage with some of the EN's features through the Spotify API. For example through a small program on the website.⁷⁸

5.2.2 My purpose

In this chapter, I am going to explore and discuss how music analysis can be practiced with Echo Nest metrics. I will discuss the prospects and raise methodological awareness of applying machine-learned metrics for music analytical purposes. As I am investigating large-scale analyses, I am especially interested in revealing consistent biases that might distort the results on macro level. But the underlying premise for this study and discussion is that the better you understand basic mechanisms on micro level, the better equipped you will be to interpret the results on the macro level. So even though I search for methodologies for investigating music from a macro perspective, I will look more closely the Echo Nest API on a zoomed in level.

Thus my aim is to achieve *general* rather than detailed knowledge about Echo Nest's features and the mechanisms behind them. Close investigation with Echo Nest's features would quickly become obsolete with the advent of the next analyzer version, but theoretical concerns persist longer. Therefore my goal here is not to engage with Echo Nest's features systematically, but rather to pick examples that illuminate generally applicable issues. Though I have engaged with features derived from a wider scope of artists, this chapter will be mostly centered around the features of all songs by Björk and Radiohead⁷⁹. These were retrieved through the Echo Nest API, May 2015.

⁷⁷ the.echonest.com, retrieved June, 2015. In June 2017, there were 38 million songs.

⁷⁸ <https://developer.spotify.com/web-api/get-audio-features/>. Or for many tracks: <https://developer.spotify.com/web-api/console/get-several-tracks/>, retrieved January 23, 2017.

⁷⁹ Remixes are excluded.

5.3 The Features

5.3.1 Echo Nest's features

The EN features can be split into two sub-categories: One category that contains the born-analogue features which comprises musicologically well-defined features, measures⁸⁰ used in traditional musicology and now computed into algorithms. The other category contains features defined by Echo Nest. These features do not represent traditional measures in musicology.

The musicological well defined features

The musicological well-defined features are Tempo, Key, Mode (i.e. either major or minor)⁸¹, Time Signature, Loudness, and Duration. As an addition to each of the Tempo, Key, Mode and Time Signature values, the Echo Nest API provides a Confidence Value, ranging from 0 to 1. This value indicates how sure the Echo Nest Analyzer is that the coherent value is the correct number.

Echo Nest defined features

The EN-defined features include the following features: Energy, Liveness, Speechiness, Acousticness, Danceability and Valence. These are examples of high-level features, created by combining lower-level ones. Despite not being well defined in a traditional music analytic sense, I presume that these features are well-defined in the sense that the algorithms calculating them are consistent, as long as the same analyzer version has calculated them.

Tracks and songs

The Echo Nest API discerns between songs and tracks. Track is each digital representations of a song on individual streaming service in the Echo Nest API; such as Spotify, Deezer, Rhapsody, 7digital, etc. Madonna's *Like a Virgin* from the *Like a Virgin* album on Spotify is, for example, a separate track from Deezer's *Like a Virgin*, but both these *tracks* are hierarchically situated under the overarching category "song" *Like a Virgin* by Madonna. Also, there are several songs named *Like a Virgin* by Madonna; one for each separate recording that contains the song.

⁸⁰ Throughout this thesis, I apply the words *measure* and *metric* in the statistic sense. See also Section 1.7, terminology.

⁸¹ There are a lot of other tonalities, but Echo Nest only ascribed music one of these two tonalities. I do not know how this is estimated, but I do know that MIRtoolbox' algorithms estimation of mode is based on probabilities of either minor or major.

For my purpose, I applied the standard way of collecting the most important and manageable of the features. I extracted the “audio_summary” features for each song and track, implying that I had one vector for each track and song containing one value for each of the features. Creating one vector per song is good if you want to calculate similarities, classify songs, or compare overall characteristics of songs (George Tzanetakis 2014, 7:00).

5.3.2 The aim with the features

The purpose of Echo Nest’s features is not exact precision. Rather their aim is to “represent the aboutness⁸² of the song in single floating point scalars”, as Echo Nest co-founder⁸³ Brian Whitman explains in a blog post (2013, my footnote). In other words, the features are designed to provide an approximate overview of a song. It was, for example, of practical rather than analytic reason that Echo Nest introduced the Speechiness feature, as they wanted to be able to automatically discern between whether a track consists of music or speech, such as in an interview with the artist (The Echo Nest blog 2011). Additionally, the generation of features ultimately has commercial and not music analytical purposes. But as Marsden e.g. points out this does not necessarily render them useless for music analytical purposes (2009, 146). In other fields of the humanities researchers have acknowledged the potential of commercial datasets, and have used them for conducting large-scale analysis (Rogers 2013, 21).

Notwithstanding, the purpose influences the approach to feature generation. The Echo Nest employee, Brian Whitman’s choice of the word *aboutness* is a strong indicator of this way of thinking. In his blog post on Echo Nest’s methods, Whitman also mentions that a lot of time is spent on quality assurance, implying a lot of practical engagement with the methods (2013). Thus, the features are created in a recursive process between machines and humans. This approach influences how the features are created and what they measure.

“These attributes are either heuristically or statistically observed from large testbeds: we work with musicians to label large swaths of ground truth audio against which to test and evaluate our models. Our audio analysis can be seen as an automated lead sheet or a computationally understandable overview of the song: how fast it is, how loud it gets, what instruments are in it.” (Whitman 2013)

⁸² Not to be confused with the Library and Information Science concept *aboutness*. Whitman’s application of the word implies approximations of a song’s main characteristics in terms of numbers.

⁸³ According to <https://notes.variogr.am/about/>, retrieved June 1, 2017.

Words like *heuristically*, *lead sheet* and *overview* are good indicators of the approach and goals.

Another Echo Nest employee, Jason Sundram, explains in another blog post that many of their features are machine-learned:

“Our attributes depend on ground truth data generated by The Echo Nest’s awesome Data QA Team, a passionate group of musicians and music lovers that includes several Berklee students. When they tell us a song is danceable, we believe it.” (2010)

And the Echo Nest website explains further:

“There’s a whole range of low and high-level features that come into play with these. They are self-selected and weighted according to a training set of labelled songs and a non-linear model. So it is not easy to explain precisely how each of the songs is being estimated.”⁸⁴

But even though the features are created heuristically, and their purpose is to provide overviews, the features represent certain musical aspects derived from the sound signal. And these certain aspects are the ones that are found reflected in the feature values.

5.3.3 An example: Danceability

One illustrative example of the approach and the outcome is the Danceability feature. A song’s danceability is a combination of various cultural, subjective, acoustic and other factors. Consequently, some would probably claim that it is ruthless to even try to measure danceability from sheer measuring on the audio files, as it is so subjective. Echo Nest are also aware of this fact, and one of their employees, Jason Sundry, state that “[w]e each groove to different music; what constitutes dance music is inherently subjective.” (Sundram 2010)

But with Echo Nest’s purpose in mind, calculating danceability has the purpose of being able to distinguish some types of music from other types. And the music people call danceable tends to have acoustical and musical aspects in common. These are sought calculated into the model:

⁸⁴ developer.echonest.com/forums/thread/351, retrieved October 23, 2014

“We use a mix of features to compute danceability, including beat strength, tempo stability, overall tempo, and more. One cool thing that I’ve noticed is that remixes of songs tend to have a higher danceability score than the originals.” (Sundram 2010)

One peculiar outcome of this approach is that in March 2014 the song with the highest danceability was the sound of a ticking clock. Probably because it holds the characteristics mentioned above: It has a very stable and danceable tempo of 120 BPM, a good tempo for dance music. The beats (i.e. the *tick tack*s) are very prominent in the mix (in fact, they are the only thing there). Normally, the sound of a ticking clock is not considered a dancefloor filler, and this example demonstrates that even though the EN feature labels describe qualities that ordinary people use to describe music, the methods are not able to measure these qualities precisely. The goal is approximations. And Echo Nest does not either claim to have found the formula for calculating danceability.

5.3.4 Transparency

The line, “[s]o it is not easy to explain precisely how each of the songs is being estimated” from the quote above is important to remember, if you are using the features for music analytic purposes. This uncertainty, expressed even by developers, is caused by the complex machine-learned algorithms that constitute the basis for the algorithms. For Echo Nest (and engineers by and large) the intermediate results are not interesting in relation to the overall purpose, and maybe therefore they have not released the precise formula for these measures. This result is that you can never be sure precisely what they measure. Even if you knew the algorithms, they would probably be so complex that it would be practically impossible to translate it into music analytic value.

5.4 Introduction to the Values - the Basics

5.4.1 First impression: They work according to their purpose

I collected Echo Nest’s features for some artists of western popular music⁸⁵. In Figure 9, I have compared the features from three songs⁸⁶. The algorithms are able to create fairly

85 The datasets with features are uploaded here: <https://zenodo.org/record/803950>

reliable information; there is generally a reasonable correlation between Echo Nest values and how the music sounds.

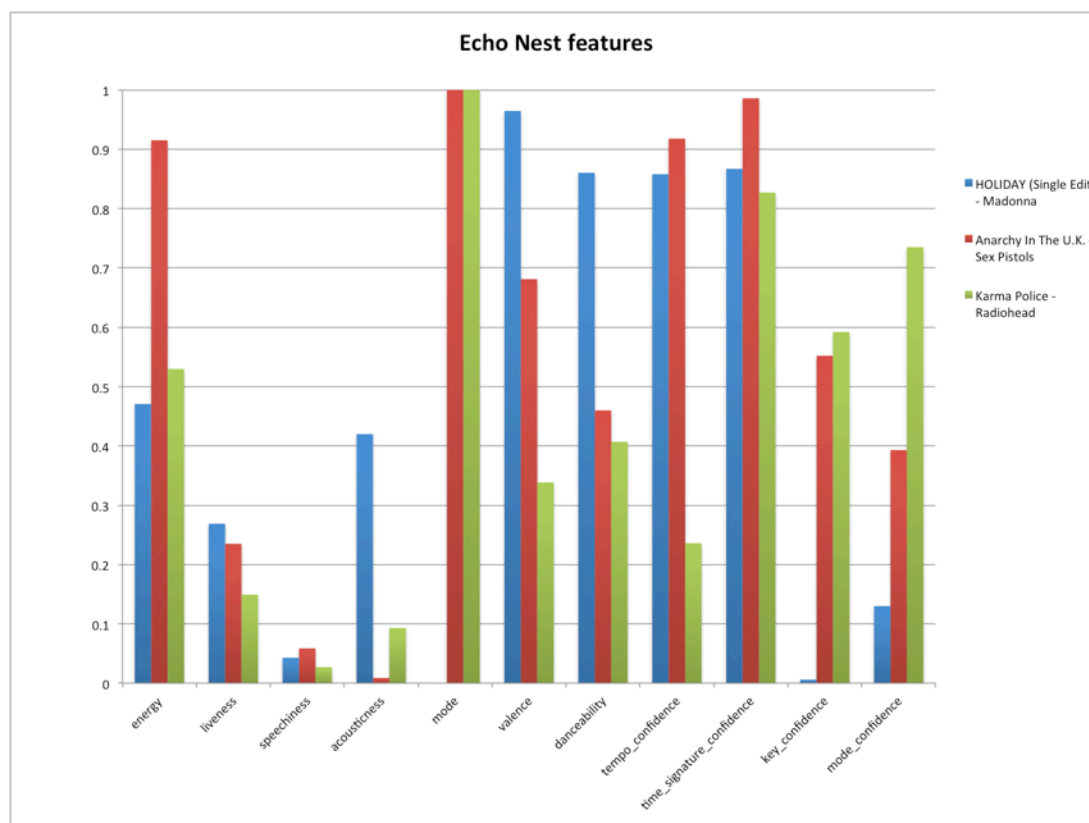


Figure 9 Echo Nest's audio summary features for three songs. mode 0 implies minor, 1 major. The tempo values are 116, 127 and 75. Time signature is 4 (meaning 4/4). And keys are estimated as A minor, E major⁸⁷ and G major (for Madonna, Sex Pistols and Radiohead).

The majority of the features agree with how the music sounds: The Tempo and Time Signature values provide satisfactory estimations, and the Echo Nest defined features largely depicts how the music sounds; Anarchy in the U.K. is the most energetic of the three songs, Madonna's Holiday is estimated as the happiest and most danceable. However, there are also features that correspond less well with the music, for example, Liveness and Acousticness and Tonality. This level of precision is somewhat characteristic for feature calculation, but rather than commenting on individual mismatches; I will point towards more general principles and concerns interpreting the features, and machine-learned features in general.

⁸⁶ The Echo Nest API holds many versions of each of the songs in the diagram. The versions visualized here have id's: SOKFZQA13777B059F5, SOBNJRV13FAE88DA93, SOHJOLH12A6310DFE5.

⁸⁷ The keys of the 16 different versions of Anarchy in the U.K vary. Within the dataset, 7 of the songs are ascribed G-major, 6 H major, 1 C major, 1 E major and 1 G# minor. This is due to different tonalities in different recordings.

5.4.2 Same song, different values

Even though there are no noticeable and hearable differences between two tracks belonging to the same song, they still have different values in the Echo Nest-defined values. Figure 10 shows a diagram, which demonstrates that different variations of the same songs often are attributed different energy-values.

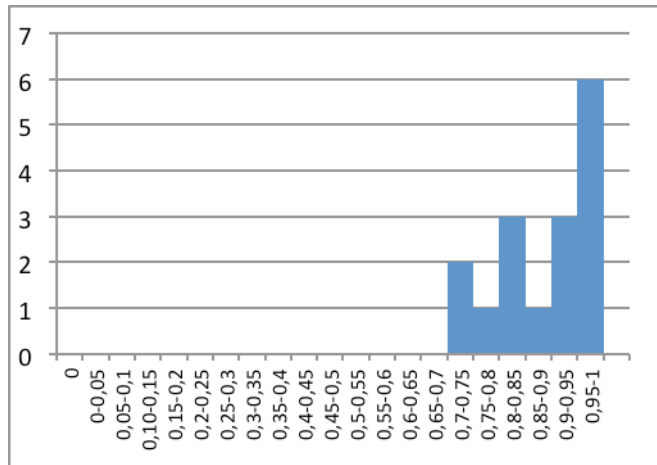


Figure 10 Distribution of energy of the 16 different versions of *Sex Pistols Anarchy* in the U.K. in the dataset. The y-axis represents intervals of energy. The height of bars display how many songs that hold the corresponding energy value.

Amongst the musicological well-defined features (except loudness) differences are less frequent, but they do occur, even within two tracks belonging to the same song.

There are several possible reasons for the variations of values:

- Different streaming services encode the audio file in a different format, implying that the acoustic signal of the same song varies.
- Different tracks within the same song may be analyzed by a different version of the Echo Nest Analyzer.
- Two almost identical songs might not stem from the same album, and hence they can be mastered differently.
- The Echo Nest metadata could be wrong.

5.4.3 Is musical progression calculated into the model?

Most songs are not constant, and to assign only one value for each feature is a coarse reduction. Though this reduction may be useful for various purposes, the way the individual parts are summarized down to one number plays a role. Misconception can, for example, occur if values are calculated from averaging feature values. A song which contrasts low and high energy passages will most likely be perceived more energetic than its average value because the contrast itself creates tension and energy. *Creep* by Radiohead has an energy value 0.35⁸⁸, which is relatively low when you take the loud energy outbursts in the chorus into account. Oppositely a monotonous song will in many cases be perceived as less energetic than the average of its individual parts. The fact that the sound of a ticking clock held the highest danceability-value is a further indicator that musical development may not be calculated into the model.

5.4.4 Are subtleties?

Another point to be aware of when interpreting machine-learned metrics is to what degree Echo Nest calculate subtle musical traits into the models. A single note out of tune or out of scale, or a displaced rhythmical accentuation can have a large musical effect, but it is uncertain to what degree such factors are calculated into the model. Radiohead's *Fitter-happier*, a track consisting of a sole computer voice speaking, does not stand out on any of the EN-features. Echo Nest ascribe the song a danceability value of 0.48, which is a further indicator that the danceability metric is very concerned with beat prominence, and loud percussive-like sounds. Hence, is more concerned with more overall properties of the sound, and less concerned with the precise placements of accentuations.

5.4.5 The features are not perfect

As I wrote in Section 4.5.3, there is a glass ceiling of about 70-85% correctness for MIR tasks (W. Bas de Haas and Wiering 2010; Li, Ogihara, and Tzanetakis 2012). Echo Nest's calculations of Key, Mode, Tempo and Time Signature are not perfect either. In the case from above of Madonna's *Holiday*, the computer mistakenly guesses that the key is A minor. But the key is decidedly not A minor, and that is probably also why the key and mode features for the song are flanked by low Mode and Key Confidence Values. Even for tracks within the same song, tonal and metric features can differ, which indicate that the audio compression format can influence even these features. If Echo Nest ascribe the same

⁸⁸ The average of three different "songs" in the database.

song different features, these differences are not continuous as the Echo Nest-defined features: If the key is mistaken, it is most often for another key within the same tonality, the parallel, subdominant or dominant key. If the tempo is wrong, it is most commonly for the double or half of the perceived tempo.

5.4.6 Are Confidence Values useful?

Echo Nest employees write that “[c]onfidence indicates the reliability of its corresponding attribute. Elements carrying a small confidence value should be considered speculative” (Jehan and DesRoches 2014). In their essence Confidence Values are features about other features. They are presumably also the features that are most complex to account for mathematically and how they correlate with musical aspects. Nevertheless, they may have music analytic value, as one could speculate that they may indicate whether there are aspects in the music that do not fit to either common or machine detectable schemes. In Chapter 7, I will demonstrate how I took advantage of the algorithms being uncertainty; when the algorithms suggests that the tempo shifts a lot, it indicate complex rhythms, such as syncopations. For now, I will leave it as an open question whether there is correlation between Echo Nest’s Confidence Values and music complexities, though I have seen some indicators of this.

5.5 Epistemological Value – How to Interpret

But how do these factors affect the analyses one can produce with the Echo Nest features? In order to answer this question, I will elaborate on two aspects of it; the epistemological and the practical value of the methods. The first (covered here in 5.5) will discuss what machine-learned methods measure, the second (5.6) will provide examples of how you despite the uncertainties and complexities covered in 5.5 nonetheless can apply the features for creating music analytical value.

5.5.1 Echo Nest’s own analysis

The question that concern the epistemological value will take its point-of-departure in a couple of large-scale graphs of many of the features’ development within the 5.000 *hotttest*⁸⁹

⁸⁹ Yes, with 3 t’s. Hottnesss “corresponds to how much buzz the artist is getting right now. This is derived from many sources, including mentions on the web, mentions in music blogs, music reviews, play counts, etc.” (Lamere 2009)

songs from each year from 1950-2013 (Echo Nest 2013b). These graphs were covered by the British newspaper the Guardian in November 2013, which treated these figures as facts, as if they were able to answer some of the most fundamental questions about how we relate music and some of the most common claims: “Why is it all so loud? You can't dance to this, not like in my day” (Dredge 2013). But one pertinent question that arises in continuation of the analyses is what we can learn about the development of popular music from reading these graphs.

Echo Nest's analyses are not scientifically reported. And the methodology behind them is consequently not explicated particularly well. Perhaps because of Echo Nest's *raison d'être* (i.e. commercial), the graphs are accompanied by extensive promises about the methods' proficiencies. Examples include statements such as “[i]t's no easy feat to have a computer listen to a song in three seconds and determine its emotional valence, but we've figured out how to do it” (Echo Nest 2013a). Computer scientists may be able to understand the degree of validity of such claims, but most scholars outside data scientific fields are perhaps left in doubt without sufficient knowledge on what the results imply. Consequently, one apparent problem that arises is that neither journalists nor musicologists know how to comment on them and critique them because neither are accustomed to interpreting machine-learned features. A second reason why they are relevant to cover here is that if musicologists in the future will apply machine-learned methods (and there are reasons why they should), they should be aware of the underlying mechanisms behind the measurements, and which epistemological claims that can be defended. In the context of large-scale analyses, it is relevant to consider to what degree we can live with error.

5.5.2 The tempo problem

The tempo calculations exemplify well some of the concerns that arise with audio content analysis methods, even on a relatively objective and well-defined measure as tempo. It is noteworthy that Echo Nest's digital tempo analysis largely disagrees with Schellenberg and von Scheve's manual tempo analysis (2012). The latter analysis suggests that the average tempo of hit music has fallen since 1960, based on humans listening to 1,000 top 40 tracks. Echo Nest's analysis suggests the opposite; that the tempo has risen based on automatic analysis of 5,000 popular songs from each year.

Both corpora are large, and it is consequently difficult to precisely pinpoint the factors behind the dissimilarities. But they could be caused by differences in corpora; the Echo

Nest corpus consists of 5,000 songs from each year that are the most popular *today*, and thus to some degree is based on the music that has survived the test of times. While the Schellenberg and von Scheve analysis is based on the most popular songs in *the year in question*, but only 40 songs per year. Echo Nest gauges that the average tempo of the late sixties is around 103, in comparison Schellenberg and von Scheve measures 116. If corpus differences cause these differences, it may give rise to new and interesting questions about how we regard the past's music from the view of today, compared to what was in the charts at the given point in time.

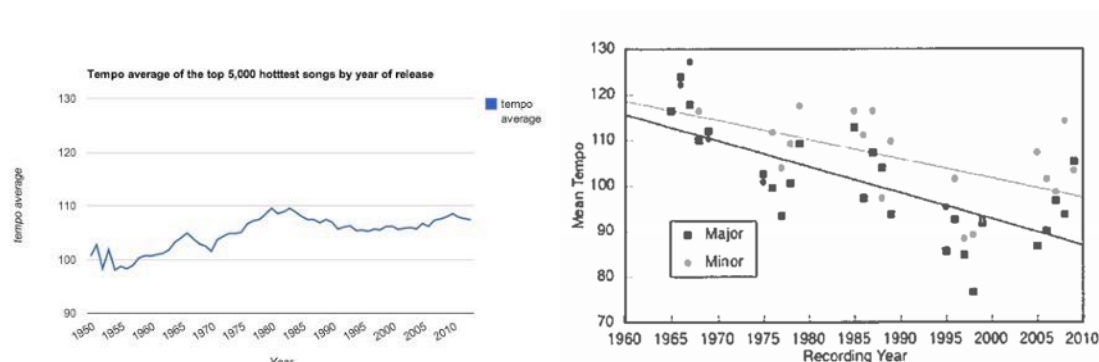


Figure 11 Tempo graphs from Echo Nest's (2013a) and Schellenberg and von Scheve's (2012, 200, Figure 1) estimation of the average tempo of Western popular music.

Some may be entitled to think that more is better (as proposed by Mayer-Schönberger and Cukier 2013); that Echo Nest must be more right than the manually because they investigate a larger corpus. But the methods play a role here, especially because MIR methods are not accurate, or at least not measuring the same as the humans do. On the one hand, when we conduct analyses of large amounts of data we can accept errors or imprecisions because they will cancel each other out on a large scale. But on the other hand, one should be aware of the risk of doing systematical biases that stir all the results.

An example of a systematic bias I encountered is the hip hop tempo bias. One example of a problematic tempo calculation is Baauer's *Dum Dum*, a song that alternates between two tempi; the one twice as fast as the other. Echo Nest's algorithm estimates the tempo to be 155. The snap, however, which enters at minute 0.30 has the function of a snare drum occurring at the 2nd and 4th beat, and due to common conventions of placements of snare drums, it will most likely make most listeners perceive the tempo as 77, which is the half of Echo Nest's estimation. This double tempo problem has occurred quite a few times during my analyses. If the Echo Nest tempo algorithm has a general bias towards doubling the tempo of hip hop beats, the error will become larger as the share of hip hop beats in the

corpus becomes larger. This bias could account for some of the disagreements between the two analyses.

But even if all tempo values were 100% correct we should still bare in mind that higher tempo means higher tempo. Period. It does not necessarily imply more energy, more power, more intensity, nor more danceability. Try comparing *Calgary* by Bon Iver with Kanye West's *All of the Light*. You will most likely experience that the former is the fastest, but the latter is the most energetic.

5.5.3 The complexity of machine-learned features

The majority of the other features⁹⁰ that Echo Nest plotted against time in their graphs are less well known in music analytical contexts. I have already sketched out some of the underlying mechanisms and mathematical operations behind them. But in the context of large-scale analysis, when attempting to understand what these graphs learn us about the music, they ought to give rise to music analytic concerns:

As I stated in the introduction to this chapter, the obvious benefit from these techniques is that computers can measure intuitively understandable aspects of the music. This entails high interpretational value on an apparent level. At first sight, you might be inclined to think that with present day's techniques - huge datasets in combination with advanced computational techniques – we actually now can measure the subjective aspects of the music.

However, we have to take the aspects, which I have already accounted for in Section 4.7 and priorly in this chapter, into account: The attention has to be pointed towards the ground truth dataset when attempting to understand the music analytic implications, even though the methodology is not accessible. A relevant question to ask is what parameters are built into the model? As I explained in 5.4, I do not have access to detailed knowledge about this. Another pertinent question is *who*? are the “music experts” who have annotated the music? and *which* music that has established the foundation for ground truth? If the algorithms, for example, had found out that music from the eighties was the most danceable, one could suspect that people who were young in the eighties had created the ground truth data. Therefore, machine-learned metrics may say just as much about those who created ground truth as about the music. As Wiering explains:

⁹⁰ Valence, organicness, bounciness, mechanism, acousticness, mode, loudness, danceability, tempo, energy

“[T]he ground truths that are used [...] are the result of very complex processes, in which perception, personal background, training and taste are at least as important as the musical content that is being judged.” (2009, 1)

5.5.4 Black boxing

A central concept to understand this general is what Rieder and Röhle amongst others label *black boxing*. They explain:

“Paradoxically, the practical need to formalise contents and practices into data structures, algorithms, modes of representation, and possibilities for interaction does not necessarily render the methodological procedures more transparent. Transparency, in this case, simply means our ability to understand the method, to see how it works, which assumptions it is built on, to reproduce it, and to criticise it. Despite the fact that writing software forces us to make things explicit by laying them out in computer code, ‘readability’ is by no means guaranteed.” (2012, 76)

Rieder & Röhle refer specifically to machine-learned techniques:

“Many of the techniques issued, for example, from the field of machine learning show a capacity to produce outputs that are not only unanticipated but also very difficult for a human being to intellectually reconnect to the inputs. Despite being fully explicit, the method becomes opaque.” (2012, 77)

Thus, even if we accept the underlying premise that we actually *can* measure subjective qualities to some extent – a premise I find sound. And even if the individual estimations of the subjective qualities do not match arbitrarily with the music, we are still far from being able to actually measure these qualities. Most importantly, it is hard to grasp the music analytic consequences of measuring music this way. Not only because a large number of musical components are brought into play in each feature, but also because of the mathematical operations’ link to these in non-transparent ways. I have created Figure 12 to display that a lot of different aspects that have to be taken into account when we want to understand how machine-learned features relate to the music.

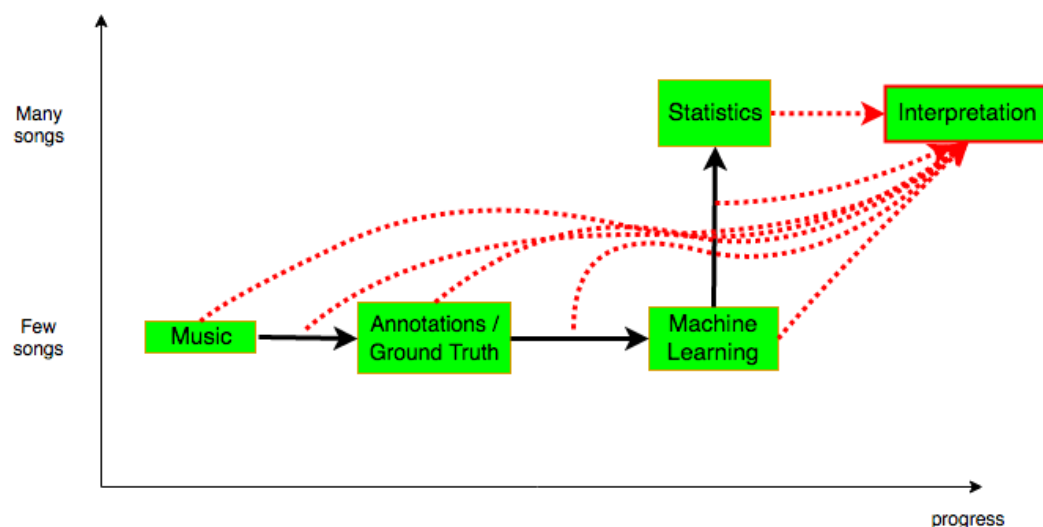


Figure 12 Flowchart of machine-learned features. The black arrows indicate the flow chart of Echo Nest's statistical analyses. The red arrows indicate that if you want to interpret the statistics you should take all components as well as transitions into account.

5.5.5 How to approach the measures

At the same time we should also remind ourselves that "[m]any MIR systems appear as if they were "listening" to the music when they are actually just exploiting characteristics of the music", as Sturm and Collins (2014, 1) have expressed it. We can deploy this knowledge when attempting to figure out what the measures imply. And one way of dealing with a part of this complexity problem is to investigate how features and music characteristics relate to each other.⁹¹

With the Danceability-feature, Echo Nest seeks to exploit that there are general correlations between certain characteristics of a song and how much it invites to dance. They aim to build a model that takes these characteristics into account for estimating songs' danceability. However, models are approximations, limited by the parameters built into them. So if you want to learn how danceability-values correlate with musical characteristics, you need to investigate the methodology and compare features with the music, preferably in combination with knowledge about the algorithms.

If I for example visualize all Björk's songs according to their danceability, I can approach this connection between features and musical characteristics (see Figure 13). Interestingly, the algorithms estimate Cocoon to be the most danceable song. After having investigated several other songs' connection between danceability-estimations and musical

⁹¹ In the Guardian article there are a few explanations of how musical aspects correlate with given features.

characteristics, my best guess is that this song scores high Danceability because of pregnant percussive elements (high pitched click sounds) and a steady tempo; these are characteristics that often correlate with danceable songs. However, the algorithms seemingly do not take the lack of drums into account. This is also the case for *Solstice* (Danceability 0.74), which does not even contain percussion, but only plucked string sounds with short attacks playing on every quarter note. *Solstice* is not either the prototype of a dancefloor-filler, more likely the opposite. Both *Cocoon* and *Virus*, which also score high Danceability, are further examples of the double tempo bias, which probably also inflicts a higher Danceability value than otherwise. The case corresponds to the double tempo example discussed in 5.5.1: Echo Nest estimates the tempo at 120 BPM while there are musical components that suggest that the tempo more likely will be perceived as the half. This discrepancy possibly influences the songs' Danceability value heavily, simply because 120 BPM in most cases are much more danceable than 60 BPM. In contrast, songs as *Bilavisur* or *I Dansi Með Per* are more danceable than their danceability value, but these songs' percussive characteristics are not so prominent in the mix. This exemplifies that MIR algorithms generally are sensitive to mastering.

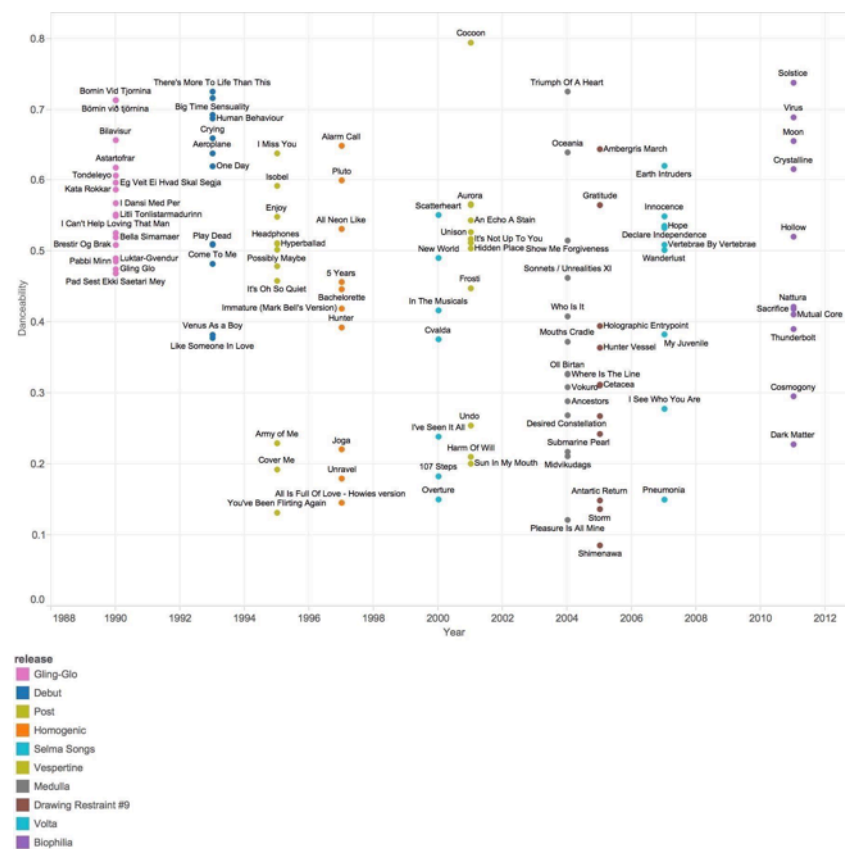


Figure 13 Björk's songs 1990-2011 arranged according to danceability and release year of the album on which they were released.

The Acousticness-features expose another example of a discrepancy between the feature label and the music's characteristics. The EN-algorithms ascribe Madonna's *Holiday* highest acousticness-value of the three songs displayed in Figure 9 above, although it contains almost entirely electronic instruments. However, the sound of *Holiday* is softer, less distorted and less noisy than the other two songs, and that is probably why the algorithms interpret it as more acoustic. Correspondingly for the Radiohead's albums, *Kid A* has the highest average Acousticness-level⁹², despite containing mostly electronic sounds. But again the overall sound of *Kid A* is more mellow and less distorted than their other albums; *Kid A* contains a lot of sonorities that have more in common with acoustic sounds than distorted guitars have.

These two examples serve to demonstrate a more general message about machine-learned metrics: Music analytical thinking has to be activated if you want to understand what the calculations imply. To understand each of the qualities that Echo Nest claims to be able to measure you have to be able to connect music and features at a deeper music analytic level. And you will end up concluding that the relationship between feature and music is very complicated. The metrics say as much about the musical characteristics behind the methods as they do about the quality modeled. So, until best practices regarding the use of a particular automatic method has been established, we have to conduct this double interpretation, I have accounted for above: We have to interpret the connection between music and feature calculations and thereafter we can interpret summarizing statistics. Machine-learned techniques do not eliminate human interpretation, they complicate it, and enlarge the scope of elements that has to be taken into considerations.

5.6 Practical Value - The Usefulness of Reductions

5.6.1 Basic statistics

On the other hand, many measures in other scientific fields are very complex but still useful. Think for example of GDP, a measure which in many regards is very problematic and complex to account for, but nevertheless is a good and practical indicator of many other aspects of a country. Echo Nest features are not designed to produce music analytic insight, but to quantize and represent *aboutnesses* of songs; to be able to distinguish songs

⁹² Average Acousticness-values for Radiohead albums: Kid A: 0.51, The King Of Limbs: 0.45, In Rainbows: 0.43, Amnesiac: 0.42, Hail To The Thief: 0.34, OK Computer: 0.21, The Bends: 0.17, Pablo Honey: 0.01

from each other by assigning them a few numbers. Therefore, the metrics are designed to cover the most fundamental aspects of how we categorize and group music according to its sound. And the metrics, therefore, enable very general descriptions of music. You win something in generality and being able to overview large corpora, but the price is that you lose something in precision and level of detail. But because of the potential immense sizes of today's datasets, you also win something in the potential to include and compare many different songs. The reduction of musical qualities down to one number can become a strength for some purposes, as long as we have a good dataset. Especially in combination with the ease of visualizing provided by modern software.

It becomes possible, for example, to plot the development in features in Björk's songs through the years to get a rough overview of the development of her oeuvre:

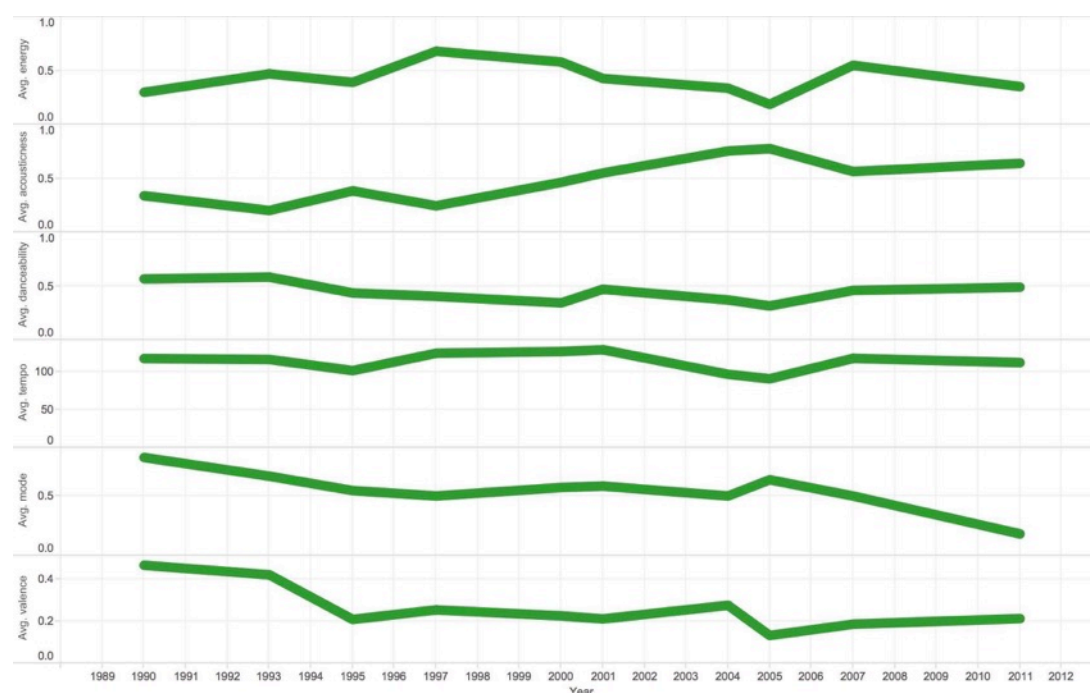


Figure 14 The development of Björk's oeuvre measured in Echo Nest features.

Graphs like this provide a coarse overview of tendencies throughout the years: They indicate that her songs have become less energetic, ascribed higher Acousticness-values (probably implying that they are less distorted) and more songs are assessed minor. We can also see that the album *Drawing Restraint 9* from 2005 stands out on several parameters.

Additionally, it is easy to compare the features of all her songs with Radiohead's songs:

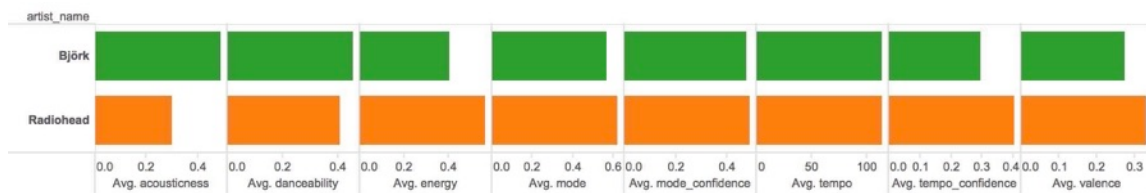


Figure 15 Average features of Björk, compared to Radiohead.

Echo Nest estimates Björk's songs as being less energetic and more Acousticness-tic than Radiohead's. In addition, the Tempo Confidence Value indicate that the algorithms have more difficulties calculating tempo from Björk's songs than Radiohead's. This could be a sign that the rhythms in Björk do not follow ordinary schemes as much as Radiohead's rhythms.

It is also possible to create a quick comparison of the two artists' songs from different periods:

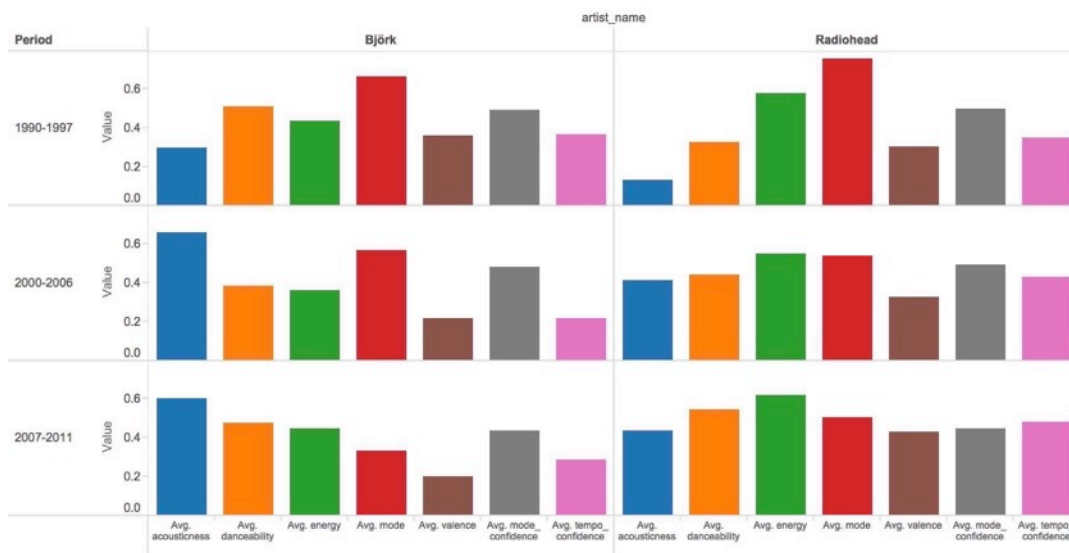


Figure 16 The development of Björk's and Radiohead's features.

These graph indicate that the sound idioms have changed for both artists, towards more Acousticness. The algorithms estimate Björk's music to have decreased in valence, in contrast to Radiohead's music. This could be connected to Björk's enhanced musical complexity, i.e. the difficulty for the algorithms to estimate tempo and mode, and that the algorithms estimate that she writes more songs in minor.

The major benefit with the digital approach is that these graphs can be created very quickly and that they contain a lot of condensed information, dependent on where we

direct our attention. The algorithms can also help us perform coarse music analysis of many pieces of music, help us where to focus the next investigation, so we do not have to listen to hundreds or thousands of songs. The graphs do not confirm anything that is not very complicated to account for, but we can use them for developing or strengthening our arguments.

5.6.2 Mapping songs

One statistical disadvantage of the graphs presented above is that averages do not account for individual differences. A data analysis technique that can help overcome this problem, and which I find suitable for exploring music, is multi-dimensional scaling, which practically scales many dimensions down to few while preserving as much information as possible.⁹³ It makes it possible to map many pieces of music to one map as the large map on everynoise.com⁹⁴ of more than 1,000 music genres demonstrates finely. The principal component analysis (PCA) (Bro and Smilde 2014) is such a technique, which can “squeeze” a high dimensional feature space down to fewer dimensions. This scaling down can make the elements in the dataset visualizable in one plot, and create an even cruder, less detailed view of the pieces of music. The clear benefit of the technique is that it enables you to detect larger patterns in the dataset, and these patterns can be found both regarding songs and features. PCA is a way to learn from the data, where you do not need to know everything beforehand. You do not need a hypothesis beforehand; all the information is in the dataset.

Figure 17 displays a PCA plot of all songs by Radiohead and Björk in the Echo Nest API. Below the “map” is a plot of the loadings, which indicate how the features relate to each other and the map. The loadings’ plot illustrates a tendency that high Valence correlates with high Energy and Danceability. In contrast, these properties correlate negatively with acousticness.⁹⁵ The plot of the loadings also indicates how the features transfer to the map above. Consequently, the PCA plot shows some of the same tendencies as delineated in the charts above. There are more green marks in the left side of the plot. This indicates that many of Björk’s song have higher Acousticness values than Radiohead’s. There are especially many squares, which indicate that many of Björk’s song from 2000-2006 have high Acousticness, and low Energy, Danceability, Tempo Confidence, etc. You can go on exploring more details in your corpus from looking at the plot, listen to the music, re-arranging the plot, listen again, etc.

⁹³ Statistically by maintaining as much variance in the dataset as possible.

⁹⁴ Retrieved February 10, 2017. See also Section 4.4.4

⁹⁵ See Appendix 2 for detailed calculations of how much the variables correlate.

songs of 7 Danish artists.

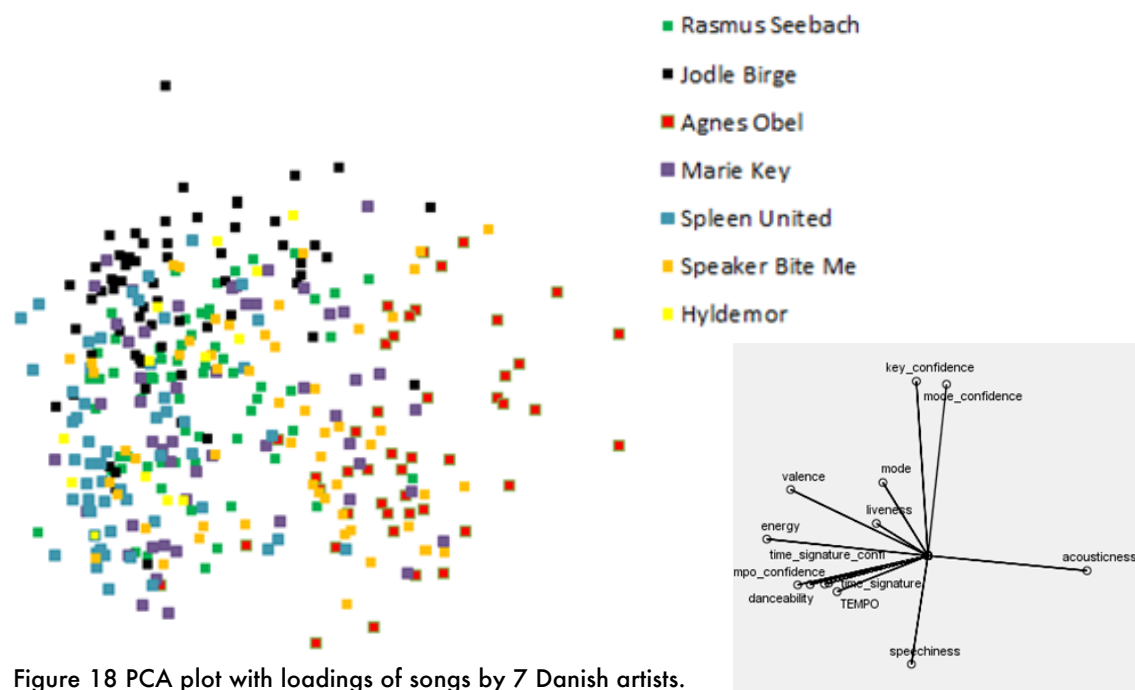


Figure 18 PCA plot with loadings of songs by 7 Danish artists.

Most importantly, I observe that the plot fits very well with the artists' styles. Interestingly also, the loadings' plot corresponds to Figure 17's⁹⁶. This indicates that the features correlate similarly, despite the differences in genres. Apparently, Echo Nest estimates the majority of songs by Danish schlager artist Jodie Birge as happier, they are more often assessed as major, and they are fairly energetic. It is easy for the algorithms to predict tonality issues on Jodie Birge's songs. In the opposite end of the map, the singer-songwriter Agnes Obel's songs presumably score low on energy, tempo, valence, and are more minor-ish.

But my choice of the word presumably also indicates a lot of information is lost in the process of reducing the dimensions with the PCA. In the Danish dataset, I arranged above, there is a general tendency that high Acousticalness correlates with low energy and valence. But we can never be sure that every dot on the map fits with the plot of the loadings. A song by Jodie Birge can, for example, have high Acousticalness, which would pull it towards the right, but it could at the same time have high Valence and Energy, and be ascribed major mode, which would make the PCA plot it to the left.

⁹⁶ It is laterally reversed, but this does not effect the interpretation of it.

The further limit of this technique is that the plot produced depends on the variable in the model. If for example only an excerpt of the variables were plotted, the plot could turn out completely different. Perhaps it would not be as useful because it would no longer be consistent with how one intuitively would expect the music to be mapped. However, Echo Nest's features are especially useful for this purpose; they are designed to provide an overview of songs, and consequently account for the most prominent aspects of songs.

Multi-dimensional scaling techniques are useful if you have a large dataset of songs that you want to map. They are well-suited for the exploratory levels of the analysis, for preliminary zooming out to the macro level. They somehow function like viewing the world from an aeroplane; you can get an idea of the contours of the landscape, and it offers insights of where to investigate your dataset more closely. But they do not confirm anything. The PCA can, for example, be applied, if you might want to create an overview of the songs played in different radio shows throughout a whole month. Or you might want to overview the development of the music of an artist that you are not familiar with. In both cases, the PCA plot can help you where to investigate your dataset or listen further.

5.7 Perspectives for Musicology

5.7.1 Which questions do models answer?

For summing up the perspectives for musicologists to apply machine-learned metrics for musicological purposes, I will commence by turning to wisdom on data analysis, alleged by John Tukey:

“Far better an approximate answer to the *right* question, which is often vague, than an exact answer to the wrong question, which can always be made precise.” Data analysis must progress by approximate answers, at best, since its knowledge of what the problem really is will at best be approximate. It would be a mistake not to face up to this fact, for by denying, we would deny ourselves the use of a great body of approximate knowledge” (1962, 13–14, quotations and italics in original)

In the case of Echo Nest, the reverse direction is equally true: Those exact answers that can be provided using Echo Nest's machine-learned features are exact answers to

questions that are not very good. Because these questions either are extremely difficult to formulate, or are rooted in data-scientific concepts with an equivocal connection to human reality. The question that accompanies the measurements of Björk's Valence values would, for example, sound something like:

If x people, aged y_1, \dots, y_x , with background z_1, \dots, z_x rated songs v_1, \dots, v_n regarding how happy they sounded, and algorithms calculated features u_1, \dots, u_m of these n songs, and an algorithm was set up to find correlations between the average rated happiness of the songs and each of these m features according to correlation model t , and this model was applied to the features u_1, \dots, u_m of Björk's songs, after Echo Nest employees had quality assured the model for s hours, what would be the development of these numbers as a function of release year of Björk's songs?

This is indeed a perplexing question. And there is a lot of variable and unknown factors that influence the results: What are x , y , z , u , n , and s for example? And how do they affect the final result? Music analytically, it becomes extremely complicated to account for the connection between features and music, because there is a large element of *black boxing* in the process from the audio file to the feature. Rieder & Röhle have expressed about this challenging relationship between machines and epistemological value that

“[o]n an epistemological plan [computers] create problems rather than resolving them. Questions of bias and subjectivity, which the computer was thought to do away with, enter anew on a less tangible plan” (2012, 73).

This also is the case for the Echo Nest-features.

5.7.2 The prospects of modeling

However, the Tukey-quote from above suggests that we can progress with approximate answers, and Echo Nest's features can provide us these. An example of such a question could be *would Americans of today find that Björk's songs have become more or less happy through the times?* The approximate answer to this question could be an indicator of the musical development in her oeuvre. The further good news is that you with digital techniques can replace Björk with any other artist and get your approximate answer instantly. We will most likely also become able to substitute happy with a broad range of other subjective or analytical qualities, such as *soothing*, *aggressive*, etc.

More generally, the model that machine learning techniques can create can become “an exploratory device, a more or less “poor substitute” for the real thing,” as McCarty (2007, 393) has remarked about models *of*⁹⁷ something. He continues: “We build such models-of because the object of study is inaccessible or intractable.” This is certainly true for Echo Nest’s features; the values represent aspects that are inaccessible otherwise: A pragmatic and heuristic way to approach the assessment of musical qualities in a vast catalogue of music. McCarty also explains that

“models-of allow the researcher to negotiate the gulf between a “limited and selective consciousness” on the one hand and “the unlimited complexity and ‘richness’ of the object on the other” (393)⁹⁸

Model building is a means of operating at more intuitive music analytical levels, a means of more directly approaching how music works in the minds of the listeners. This way of approaching music could ultimately lead to the questioning of prevailing music analytic methods.

This nature of models has to be brought into mind when applying them. Echo Nest’s features are largely in line with what they attempt to measure, but they are built on certain premises, and any interpretation of them, therefore, should take these premises into account. Rather than being a fixed set of rules, these models become “exploratory devices” that enables us to approximate answers that can become empirical components, easy to retrieve, applicable for answering other questions.

5.7.3 Current issues of modeling

If musicologists want to become better able to exploit the prospects of machine-learned methods, to be able to apply them and criticize them, there is a set of new skills that has to be acquired. An apparent challenge is that many of these methods are created on bases that stem from outside musicology’s traditional core domain. Therefore, musicologists will have to become acquainted with the basic understanding of the mechanisms behind machine-learned methods, if they want to understand what they entail. This implies knowledge about what they measure, and what they do not measure. And this implies knowledge about the type of new questions that can be posed with the methods, and what kind of answers they entail.

⁹⁷ McCarty discerns between models *of* and models *for*.

⁹⁸ Referencing (Shanin 1972, 10).

Additionally, there will be a process of accustomed to how MIR “listens” to music. What are the musical characteristics that the models are based on and thereby influences the results? On the one hand, these will vary from feature to feature. However, at the same time there might be tendencies that ACA methods induce to new predominant ways of listening at the expense of other ways. From investigating Echo Nest’s features, I see a current tendency that timbral characteristics and overall sound impressions with little regard to meaning on subtle levels play the most prominent role. Temporal aspects seem to be omitted, which de Haas & Wiering (2010) also have argued.

On the feature level, there will have to be a process of practicing with the individual methods, of becoming accustomed to what the applied feature actually measure, and how they correspond with the acoustic reality. This includes identifying pitfalls and establishing best practices. When attempting to measure a particular musical aspect there is a potential risk that the methods will become gradually improved, and therefore best practices will never become settled, which will impede comparisons across studies.

I have in the previous chapters argued that when applying ACA for music analysis, you have to interpret at multiple levels; the topic you investigate, the statistical methods, and the ACA methods (Tukey, 1962). Especially knowledge about the latter of these three will often be the soft spot. Therefore methods and results should be made available so others can contest the empirical assumptions they entail. Access to source code is preferable but may not answer all questions there are. Also, knowledge about ground truth (who created it? what kind of music is it based on?) is preferable. Additionally, access to datasets or interactive visualization options is also desirable. Empirical results and discussions of results, implications of methodology should go hand in hand.

In relation to large-scale analysis, I see the question of bias pertinent to address: You can get some sense of directions by applying methods such as Echo Nest’s. It is, for example, possible to apply advanced statistical methods, such as PCA, to structure large datasets and find patterns in it, as I demonstrated in 5.6. But in order to find out which music analytic directions the features suggest, you will need to interpret your dataset at a deeper music analytic level; detect what musical characteristics are built into the model, because the algorithms actually just measure these. However, Echo Nest’s features can be applied for music analytical studies, despite being very complicated to account for and imprecise. I regard the results as fairly reliable suggestions based not on entirely random assumptions, suitable for delineating contours in large sets of data. Focusing on all the problems the methods hold could lead to neglecting some of the prospects they hold.

5.8 Rounding off Chapter 5

In this chapter, I investigated Echo Nest's features' applicability for music analytic purposes. These features are examples of some of the recent opportunities of measuring music that has arisen with the advent of digital methods: It has become possible to apply machine learning algorithms that attempt to measure and thereby model subjective aspects of music very directly. I have discussed these features' epistemological status, arguing that it is very complex to account for what they measure. But I also argued that I see potentials in applying such metrics in practice, especially when you want to handle, organize and grasp a large dataset of songs. In the next chapter, I will investigate an example of a recent music analysis that applies other ACA techniques to investigate overall trends in popular music history. While in Chapter 7, I will practice music analysis with more flexible, less superior ACA methods.

CHAPTER 6

Then the Science Guys Entered the Room

In this chapter, I am going to discuss the epistemological value and prospects of an already existing analysis, Mauch et al.'s "Evolution of popular music: USA 1960-2010", published in 2015. This study illuminates central issues in how we can analyze music with ACA methods, and it exemplifies very well what data analysis enables us to do. The study also elucidates, how large amounts of musical data can be handled, grouped, made manageable, visualized and applied for investigating musical tendencies in thousands of songs. In Chapter 5, I argued that machine-learned features basically are compounds of lower-level features. Mauch et al. have based their study on two of the most important of these, namely MFCC's and Chroma, roughly representing timbre and tonality. Thus the examination of this study enables me to investigate closer these and how these two features translate into musical aspects, which can assist me to understand two of the most basic and promising features.

This chapter at the same time serves to be a music analytic appendix to the analysis, as I will also seek to pinpoint some of the epistemological complications that arise as a consequence of the advanced techniques. In this pursuit, I have identified discrepancies between the computer scientific approach that the study is a product of and a music analytic aim. The central question I pose, is what to do with these types of analyses, how to interpret them, and how we can exploit their advantages?

6.1 The Analysis

6.1.1 The object of their analysis

The study was set up to analyze and identify trends in tonal and timbral properties of the music that has occurred on the single chart US Billboard Hot 100 between 1960 and 2010, a total of more than 17.000 songs. The questions posed were mainly inspired by

evolutionary biology: “Has the variety of popular music increased or decreased over time? Is evolutionary change in popular music continuous or discontinuous? And, if it is discontinuous, when did the discontinuities occur?” (2). The authors applied a combination of data and music analytic reasoning to answer these questions. The study was not conducted by musicologists but rather by scholars from outside musicological departments, from data-scientific disciplines and enterprises⁹⁹, accustomed to handling and analyzing large amounts of data. Presumably because of the large corpus-size the study was exposed in the media in an extent that musicological research rarely does.

6.1.2 Findings

Through a complicated technical procedure Mauch et al. ascribed each song a distribution of 8 harmonic topics (H-topics) and eight timbral topics (T-topics), from measuring 30 seconds of it. The timbre distribution of a song could, for example, be 50% T3 (which was ascribed the labels energetic, speech, bright), 30% T1 (drums, aggressive, percussive) and 20% T5 (guitar, loud, energetic). Correspondingly, a song was ascribed values that sum to 1 for the 8 harmonic topics, containing information on chord changes. The topics were formed automatically by applying statistical clustering operations.

After having assigned distributions to all the songs in the corpus, it was possible to measure the prominence of each topic through the years. The topic data amongst other showed that the amounts of dominant 7th chords (H1) were declining through the years. The amount of minor 7th chords (H3) rose in the 70’s, and the no chords-topic (H5) increased rapidly with the beginning of the 90’s. Regarding the timbre topics, T5 (guitar, loud, energetic) peaked in late 60’s and mid 80’s and T1 (drums, aggressive, percussive) peaked around the early 90’s. (See Figure 19.)

Next, Mauch et al. applied statistical clustering methods to group the songs into 13 styles (4-5). They ascribed labels to each of the 13 styles by assigning them the most common tags from last.fm¹⁰⁰. From this operation, they were able to visualize that style 2, labeled “hip hop/rap” emerged in the late 80’s, and became the most dominant style from the early 90’s and onwards. While the “funk/blues/jazz/soul” style (style 4) declined through the years.

⁹⁹ The researchers were affiliated with:

- School of Electronic Engineering and Computer Science, Queen Mary University of London
- Division of Life Sciences, Imperial College London
- Last.fm.

¹⁰⁰ The webpage www.last.fm allows users to tag music. last.fm therefore holds a large dataset of music with corresponding tags which can be useful in many analytic purposes. For a genre analysis through last.fm see (Lieken, 2007).

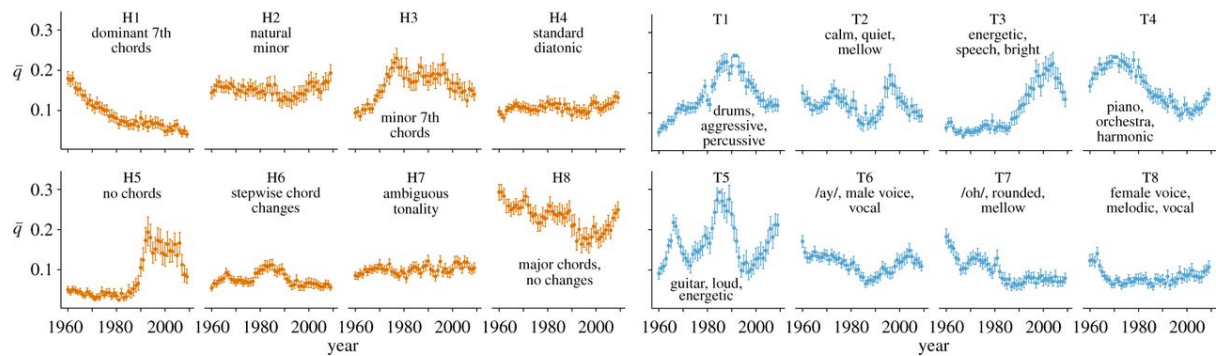


Figure 19 The Evolution of topics (Mauch et al. 2015, p.4 Figure 2).

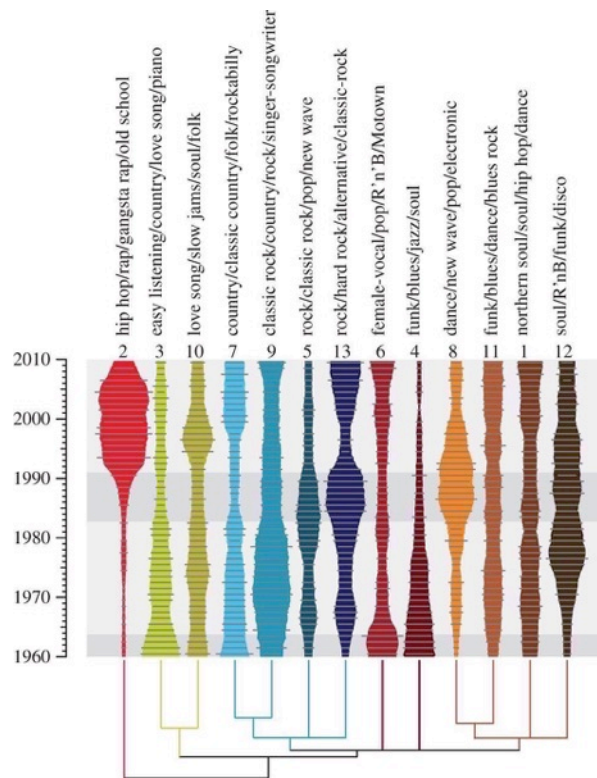


Figure 20 The Evolution of styles (Mauch et al. 2015, Figure 3)

To answer their initial questions about the change in music history, Mauch et al. measured how much the topics in one quarter of a year resemble other quarters, by measuring pairwise similarities in topics. This enabled them to identify “three revolutions: a major one around 1991 and two smaller ones around 1964 and 1983” (6). These years were attributed periods of “particularly rapid musical change” since these are the years which topic-wise resemble the adjacent years the least. For example, 1991 “is associated with the

expansion of style 2, enriched for rap-related tags, at the expense of styles 5 and 13, here enriched for rock-related tags” (6).

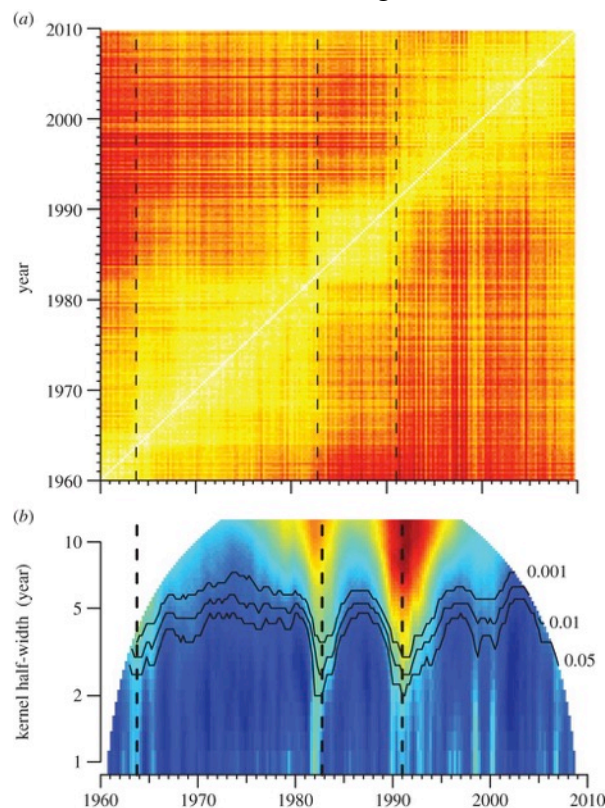


Figure 21 “Musical revolutions in the Billboard Hot 100” (2015, Figure 5).

In 3.5, the authors zoom in on the years around 1964, examining the British Invasion in the USA. I will not go into detail about this part of the study.

6.1.3 The style of the analysis

Though there is no fixed recipe for conducting big data analyses, this study accounts for a typical big data analytical approach: Large amounts of data analyzed by computers tested for patterns and correlations in various ways. Their approach can also be described as rather data-driven because data is analyzed without prior hypotheses¹⁰¹. Many both data scientists and musicologists would probably also call the study an exploratory analysis.

¹⁰¹ In practice, the relation between hypotheses and data is always more complex. One of the reasons is that “[t]ools are informed by theories about research” as Clement has explained (2012, 883). In this study, the data that is retrieved at first place is rooted in both music theory and music information retrieval theory: Tonal elements are measured because there exist a presumption that tonal properties of the music play a large role; tonal aspects have traditionally been the primary object of music analysis. Timbre is also a well-known concept in music theory, but MIR has enabled a more detailed and exact quantification via measuring MFCC’s. Likewise, the results are rooted in theories, e.g. regarding evolution.

The authors themselves end up concluding that the findings “provide a quantitative *picture* of the evolution of popular music in the USA” (9, my emphasis). A positive interpretation of this that they thereby acknowledge that this is one out of many possible ways of looking at the music at Billboard Hot 100.

Nevertheless, there are also signs that both media and the researchers themselves attribute higher epistemological value to the research conducted than to prior musicological research. I ascribe this to the application of the large corpus investigated and the methods that appear as if they eschew human interpretation. Mauch et al. do not neglect the possibility for stating that this is *real* science, unlike previous writings on music history¹⁰² which have lacked what “scientists want: rigorous tests of clear hypotheses based on quantitative data and statistics” (1). And they end up concluding “[t]hose who wish to make claims about how and when popular music changed can no longer appeal to anecdote, connoisseurship and theory unadorned by data” (9). These sort of statement are flanked by bombastic statements, which are far oversimplifying complex matters. Such as the headlines “Musical diversity has not declined” and “Musical evolution is punctuated by revolutions”. Especially the term *revolution* expresses more drastic changes than the study can account for. However, I will apply revolution myself throughout this discussion, because this study is my point of departure.

The media have well accepted the analysis and uncritically reproduced its conclusions (Akpan 2015; Thompson 2015; Gush 2015). They might be charmed by the study’s potential for creating flashing, click-baiting headlines. Or by the apparently objective and sober approach which seems finally to settle old discussions about pop music in combination with a fascination about what present-day technologies can do. But it does take a significant amount of twisting of the analysis’ findings to conclude that hip hop was a bigger musical revolution than the Beatles, that 80’s music was boring or that 1964, 1983 and 1991 were the biggest revolutions in music history.

Though it is hardly big news that the media in the presentation of this study do what they often do (present a strongly angled research results), they nevertheless deserve a remark because of the detachment between musicology and MIR. Many music researchers will probably have limited experience interpreting MFCC’s and the complicated statistical methods on which the study is based. But if they get their information through the media there is an apparent risk that it will only enlarge the misunderstandings between MIR and musicology.

102 Strangely, musicologists are not mentioned between the five mentioned groups of people writing music history.

6.2 My Purpose

In relation to the purpose of this thesis, the Mauch et al. analysis accounts for a typical example of the new types of analysis that has been made possible with the advent of ACA methods. The analysis is, therefore, a good starting point for a discussion on how to derive music analytic insights from the methods. In the following pages, I have two main aims: First, I will in Section 6.3 critically investigate the epistemological value and assumptions of the study. While I, in 6.4 and 6.5, will argue that the study actually contains valuable information, despite all the concerns explained in 6.3. In 6.4, I will outline some of the music analytic information, I derive from the study; information that I have found outside the text. While in 6.5, I will discuss how to progress from their analysis; which benefits that the generation of information and the authors' analytical approach entail.

6.3 Challenging the Epistemological Claims

boyd & Crawford have explained that “[j]ust because Big Data presents us with large quantities of data does not mean that methodological issues are no longer relevant” (2012, 668). Despite the apparent inclusion of all, the Mauch et al. analysis contains a range of choices that could have been carried out in other ways. These would have affected the results, and likely also the headlines. What would have happened if the dataset was different? If other features than the tonal or timbral were retrieved? Or if other statistical methods were applied? Any methodological choice holds the potential of affecting the conclusions profoundly.

6.3.1 The choice of dataset affects the conclusion.

boyd & Crawford's next line is the apparently paradoxical remark “[u]nderstanding sample [...], is more important now than ever” (668). The apparent paradox lies in the matter that big data techniques allow analysis of full datasets and not only samples. The study is an example of an investigation of a full dataset.¹⁰³ However, boyd & Crawford's remark illuminates the fact that no matter how big the dataset is it will always be biased. And as Mauch et al. also acknowledge in their analysis, the American Billboard Hot 100 is only a subset of all music. This complicates the generality of the claims one can infer from

¹⁰³ Though it covers only 86% of all songs that have appeared on the chart in the period (Supporting Information, 2 (M.1))

the analysis: It does not regard music or not even western popular music, rather there is a complicated relationship between the study and these two. Therefore 1964, 1983 and 1991 did not revolutionize music as stated in the text but the music on the US Billboard Hot 100, and consequently only music in the USA. Mauch et al. are well aware of this, but they nevertheless do not discuss or reflect on how the dataset biases the results, or how well the dataset accounts for popular music in general.¹⁰⁴

If Mauch et al. had conducted the analysis on the US *album* charts, 1968 might have been one of the years that Mauch et al. had found to “revolutionize music”. As Henrik Smith-Sivertsen wrote in a post on Matthias Mauch's blog: “My claim [...] is that the “revolution” would move to 1968 if you ran year test on the albums charts instead of singles.” Since “the album was originally primarily a medium for adult music. However, during the 1960's it changed radically, as pop and rock bands changed practices and attitudes.”¹⁰⁵ Furthermore, there have been varying methods of how the positions on the chart have been calculated, and this can also have biased the results. What if changes in ranking metrics have caused changes in music that caused these revolutions, Mauch et al. identified?

Detailed knowledge about what we investigate will learn us about all the *but's*. This is also why Dalton & Thatcher explain that ‘big data’ increasingly stresses “the importance of domain knowledge” (2014, #5): Despite we have automatized and data-driven analysis techniques, we can get even further with our analysis if we have knowledge about what we investigate. It would take further detail knowledge to predict the outcome if only top 10 was investigated, or top 200? And what if similar analyses were conducted on other charts? Which years would then have become the “revolutionary years”?

6.3.2 The choice of music analytic methods affects the conclusion

Secondly, both the music analytic and the statistical methods reduce music on many levels. The problem with the reduction is that it may end up distorting the results because the methods imply a lot of black boxing, which impedes the translation from numbers to music analysis.

¹⁰⁴ In Serrà et al. (2012), a comparable big data music analysis, this matter is even fuzzier. They provide an analysis of 464,411 songs, but they do not at all reflect on how these particular songs have been selected. They remark that the dataset “includes a variety of music genres such as rock, pop, hip-hop, electronic, jazz, or folk” (2012, SI p. 2), but what can we make of the results when we do not know from what they are based, and how they are biased?

¹⁰⁵ <https://theconversation.com/how-we-discovered-the-three-revolutions-of-american-pop-41624>, retrieved September 26, 2016

Mauch et al. have chosen to investigate timbre and tonality aspects of their dataset. This is a good, pragmatic choice, since tonality and timbre are considered important musical aspects, and within MIR they have been demonstrated proficient for solving different tasks. But focusing sheerly on tonal and timbral aspects only comprises a limited view of the music, and this affects all of the results and therefore relates to the generality of the findings. Mauch et al. are well aware that their study is limited regarding the objective criteria they set up: “Our measures must capture only a fraction of the phenotypic complexity of even the simplest song; other measures may give different results.” (9) Thereby they implicitly acknowledge that they have made a lot of decisions in the analysis process. But to put it bluntly, one could ask how objective the study is if they can choose the methods themselves, and thereby perhaps attain the conclusions they like? Or formulated in another way: What is the relationship between what they want to find and the way they choose to prove it?

Firstly, there is a complex relationship between the methodological choices and the results. It is difficult to tell whether the rules dictate the game, or the game dictates the rules? Style cluster 1 is for example tagged with as different genres as “northern soul/soul/hip hop/dance.” These four genres are probably grouped into the same style cluster due to how harmonies are measured and the prominent role of tonality in the study. This cluster 1 contains songs by artists as varying as Ray Charles, Temptations, Red Hot Chili Peppers, Madonna together with songs of modern pop and hip hop artists such as Nelly, Kanye West, and Ludacris¹⁰⁶. However, the songs in Cluster 1 have a high degree of H2 (minor chords) in common, and that is presumably why the clustering algorithm clutches these songs together. But the results say as much about the methods than about the music: They have chosen to measure it this specific way, and therefore the results turn out how they do. This way of grouping largely corresponds to measuring top speed, size and number of eyes of some animals and hereafter concluding that salmons are more human than ants are. But if a lot of other factors, such as social ones, were included in the measurements, it would probably have led to other results. The theoretical assumptions that guide the measurements and groupings are crucial to the outcome.

My second point regards the generality of the findings. When Mauch et al. headline their section 3.4 “Musical evolution is punctuated by revolutions” (6), it would perhaps be more precise to write something less bombastic like “timbre and tonality have changed in a non-constant pace in the songs that have appeared on the Billboard Hot 100.” Big data critics Dalton and Thatcher have proposed that “[w]e must ask what it means to be quantified in

¹⁰⁶ I have compiled a playlist with songs from these artists that adhere to style 1: <https://open.spotify.com/user/hviderusland/playlist/6slMd9hEEYNZKiZqjUxqW2>

such a manner, what possible experiences have been opened and which have been closed off?” (#2). While the quantification of music into tonalities and timbres opens for a glimpse of a larger picture of the development of popular music it at the same time closes for that the “revolutions” can be found other musical parameters such as the melodic or rhythmical. If Mauch et al. had included rhythmical factors, it might have led to other results, as Mauch et al. themselves also mention in the discussion (9). But again these questions are neither addressed nor discussed, and many musicologists may consequently find the findings presented as too conclusive and final.

6.3.3 Handling a large dataset - from complex statistic methods to music analysis

The reductions not only concern the aspects of music that Mauch et al. investigate, but they also concern practical aspects of handling the data and the use of statistic methods. There are two main aspects of managing this large corpus that blur the music analytic insights. The first is that it is hard to overview and handle the immense dataset. The second relates to the statistical methods applied in order to reduce the large amounts of data and the complexity of music to manageable sizes.

The challenges of overviewing large amounts of data are related to the working procedure: To grasp and analyze 17,000 songs, Mauch et al. have been necessitated to apply statistical music analytic techniques that implied that they probably not were completely confident of the relationship between sound and its quantification. Moreover, the statistic procedure has been automated for them to run on an amount of audio files that was impossible to overview otherwise. But there are a lot of steps in the process where errors can emerge: Two songs can become swapped, one line in a script can be deceptive, figures in the datasets can accidentally be swapped, etc.

In this study one may wonder for example wonder which algorithms make Whitney Houston’s *I’ll always love you*¹⁰⁷ be ascribed 97% T1 (drums, aggressive, percussive)¹⁰⁸, although the song does not contain an aggressive nor percussive section anywhere. If this is one single misleading number made by a wrongly digitized audio file the problem will be irrelevant since it would be subordinate to the larger picture. We can live with messy data and singular errors when the dataset gets big enough (Mayer-Schönberger, & Cukier 2013). On the other hand, it can be hard to tell whether this supposed mistake is derived

¹⁰⁷ Link to Spotify playlist <https://open.spotify.com/user/hvinderusland/playlist/37BPzD6ISA4y44Y8VzRN7I>

¹⁰⁸ https://figshare.com/articles/Main_Dataset_for_Evolution_of_Popular_Music_USA_1960_2010_/1309953, retrieved September 26, 2016

from a systematic bias that can be found somewhere hidden in the algorithms. If this is the case, it would skew all of the results to some extent. Or perhaps it is even not an error in accordance with the initial intention; the audio file is correctly measured, and distributed properly into the desired topics. However, this would impede and almost destroy the relationship between the calculated topics and their intuitive music analytical synonyms. Luckily, from scrutinizing the dataset, I have found that the other songs from the early 1990s that resemble *I Will Always Love You* are not grouped into the same style cluster 2. This indicates that this is a singular “miscalculation.”

The second point that relates to the handling of data is that Mauch et al. choose to apply machine-learned statistical techniques. These methods are suitable for managing high dimensional datasets because they reduce them to fewer dimensions, as I demonstrated in Section 5.6. Mauch et al. use these to grasp the vast amounts of data and the complexity of the parameters on which they choose to focus. But this entails that a style is formed from at least three levels of statistical reductions: Firstly, 30 seconds¹⁰⁹ from each song is measured due to timbral and tonal aspects. Secondly, these measurements are scaled down to distributions of 8 + 8 topics. And thirdly, PCA analysis was applied to these topics to assign each of the songs into one of 13 styles. The advantage is that this procedure created intuitively understandable timbre and harmonic topics with objective methods which on the one hand makes the study easy intelligible. The major disadvantage is that these reductions blur and complicate the music analytic insights.

Every time the data is reduced, we lose sight with the music’s characteristics. This is most likely a price we have to pay for being able to overview the data. However, a large element of ambiguity arises by letting the machines do the clustering; the data becomes harder to understand and new methodological questions arise. The main question we need to pose here is to what degree the boundaries between the automatically formed clusters make music analytic sense? Do they divide the music into convincing groups? What would happen if Mauch et al. had created 9 H & T-topics instead of 8? Would the 1986's synth T-topic be split into two T-topics, which would result in the 80's becoming more diverse in the analysis? Would it ultimately have resulted in other “revolutionary” years? Again, I am

¹⁰⁹ They are calculated from 30 seconds excerpts, so the values do not have to be representative of the song in its totality, but only for an unknown passage of it. Which excerpt is not obvious from reading the article and this hold yet another potential bias in the results. As Henrik Smith-Sivertsen wrote in a blog post to Mauch: “Songs tend to evolve. One of the examples used, “Bohemian Rhapsody” is built of very different blocks, and especially concerning timbre there are quite some differences throughout the song. And moving into the 1980's, intros seems to get much longer than earlier, probably in relation to the advent of MTV, making the question of where your 30 seconds are from quite important. Taking, for example, November Raion (sic!) by Guns'n'Roses (1991), it starts with orchestra and piano, the drums enter at 46 seconds, the singing voice at 1:13, and Slash really enters at 3:29.” <https://theconversation.com/how-we-discovered-the-three-revolutions-of-american-pop-41624>, retrieved September 28, 2016.

disputing the generality of the findings; if we are to make all these reservations about the dataset, and both music analytic and data analytic reductions, what does this study tell us?

Even if these methods make sense statistically, and perhaps even intuitively, it can be hard to determine the music analytic consequences of applying them. In any case, it will be impossible to conceive the entire consequences of applying them. What are e.g. the music analytic justifications of clustering songs as stylistically different as Crazy by Kenny Rogers, Young Blood by Bruce Willis, I Like to Move It by Reel 2 Real and One More Chance by Notorious B.I.G.¹¹⁰ into the same style cluster 2¹¹¹, other than objective machine-learned techniques clustered them this way? Again we can live with errors to some degree, but we should be alert when we lose the ability to overview its consequences. And once more we should ask ourselves to what extent do systematic biases occur? In the analysis, I find the link between the statistical techniques and music analytic reasoning is unclear.

Even though we are way into data-scientific territory, this is, in fact, a sensitive subject, because the data analysis is applied retrieving knowledge about a subject, music. The outcome of using the techniques becomes somewhat complicated, and it takes careful scrutinizing the dataset to comprehend their relation to the music they are applied on. This is further enhanced by the fact that best practices for conducting music analysis by measuring MFCC's are not yet established.¹¹² One may wonder whether this is a piece of complicated math or music analysis? Hopefully, it is the latter, since the argumentation takes place at a music analytic level.

6.3.4 Final remarks on the epistemological value

Mauch et al. introduce their Section “3.4 Musical evolution is punctuated by revolutions” by claiming that

“[t]he history of popular music is often seen as a succession of distinct eras, e.g. the ‘Rock Era’, separated by revolutions. Against this, some scholars have argued that musical eras and revolutions are illusory. Even among those who see discontinuities, there is little agreement about when they occurred. The problem,

¹¹⁰ Link to Spotify playlist: <https://open.spotify.com/user/hvderusland/playlist/7mpXlf3WTv6fHElQIwKOBV>

¹¹¹ Cluster 2 is labeled “hip hop/rap/gangsta rap/old school”.

¹¹² Even when I scrutinize the dataset it sometimes can be hard to translate from topic values to music characteristics because it is opaque what excerpt of the songs that is measured.

again, is that data have been scarce, and objective criteria for deciding what constitutes a break in a historical sequence scarcer yet.”(6)

It is true that this study is based on a lot data and objective methods in the sense that the results are reproducible, but this does itself not qualify the study to be less problematic and or hold higher epistemological value than previous. The study still produces a reduced views music, reduced in relation to both dataset and methods. I am not arguing against reducing music into quantities; this is necessary to analyze large amounts of music through data. Nor am I arguing against the reduction of aspects of the music you investigate; this will always be a concern, especially when producing large overviews such as these. Rather, I am arguing that the reductions should be considered and taken into account in the analysis and in the presentation of the results.

6.4 My Interpretation of the Study

The way data analysis translates to musical qualities is very fuzzy, and not guided very well by the accompanying text, which is mostly focusing on the mathematical operations. However, having scrutinized this connection between data and music closely, I will offer my thoughts on what this study indicates and what we can learn from it.

First of all, the study gives a visualizable overview of the larger picture of mainstream pop music. It visualizes the rise and fall of varying idioms somewhat convincingly.

Furthermore, it is acceptable to conclude that the study shows that something changed in chart music around 1983 and 1991. I am reluctant to argue that music technology is the single biggest driving factor for these changes. But it is also important to remember that correlation does not imply causality. Hence, I do not imply that music technology is the primary driver behind musical revolutions but only that the applied algorithms are very good at detecting changes in music technology. In the analysis, there is a general tendency that timbre topics change faster than the harmonic topics. This suggests that sound idioms change faster than which chord progressions are used when both are understood in these very broad categories. Changes in timbre consequently become a more influential factor than harmonics when comparing two years with this method. 1983 seem to stand out because of the entrance of the synthesizer and because of a revival of rock music on the charts. While 1991 primarily stands out because of a large increase in hip hop and club music entering the chart around that time. This brought along a harmonically sparser

sound, with more focus on the beat, more space between chords, and more rapping. The emergence of this sound on the Hot 100 probably influenced three topics since it brought along both a sudden high increase in H5 (no chords)¹¹³, more T1 (drums, aggressive, percussive) and T3 (energetic, speech, bright) in the early 90's.

Another point is that the detected changes do not regard music in general, but American mainstream's acceptance of the new technologies, and the new styles it brought along. These were mediated by record companies', which promoted artists that fit with current sound idioms into the market in agreement with public acceptance of certain sounds. Consequently, the statement that "[m]usical evolution is punctuated by revolutions" (6) is imprecise, since the revolutions detected do definitely not concern musical evolution alone. It would probably be more precise to write that the study deals with the music, which was accepted by the mainstream. But this is still not an adequate explanation, amongst other due to the reasons argued in section 6.3.

Methodologically, the study suggests me that MFCC's are sensitive to mixing. For example, I suppose that the gradual increase in T1 is caused by the louder mixing of the drums, especially the snare drum. I could easily imagine that a remastering of a song could affect its topic values profoundly.

The study also demonstrates that there are different ways to measure and thus argue for homogenization or not. Mauch et al. conclude that they find "no evidence for the progressive homogenization" (5). Thereby they offer another say in the empirical examination of whether popular music is getting increasingly homogeneous or not. Another prominent music big data study by Serrà et al. who found a tendency of increasing "restriction of pitch transitions" and "homogenization of the timbral palette" (2012, 1) is another prominent case in this debate. These two studies discrepancy demonstrate well that a discussion about homogenization should have its foundation in theoretic notions. Whether popular music is getting more homogeneous or not lies in the underlying theoretical assumptions that guide your methodology: For example, is the lack of tonal content (H5 - no chords) another thing or a simpler thing? To judge from Mauch et al.'s methodology, they presuppose the first, while Serrà et al. presuppose the latter. However in both cases, theoretical reflections are omitted, and the objective methods seem rather to end discussions on this notion when they, in fact, could initiate an interesting discussion. Robert Fink provides the most plausible explanation to the results: music is "getting funkier" (2013).

¹¹³ Below I will argue that no chords could be regarded as more of a timbral than a harmonic aspect.

But. I hope you have read all these suggestions for interpretations with a permanent but in the back of your head. The transparency when translating from data to music analysis is blurry, and the main problem is that the formation of topics is not self-evident; it is caused by the corpus, and theoretical assumptions behind the methods, added a very significant element of black boxing. The way the topics automatically are formed influences all of the results. And yes they are to some extent intuitive and one acceptable way of dividing music into categories, but it is not the only possible way, and one could also easily imagine other formations than these. That the algorithms were set up to detect “no chords” beforehand is not self-evident, but perhaps *the* methodological choice that influenced the conclusions the most as the biggest “revolution” first and foremost was found due to the excessive increase in the “no chords” topic. Music analytically, one could easily argue that “no chords” is more of a timbral aspect than harmonic since harmonies often are there and work in the music, they are just not drawn out; the tonal elements becomes less prominent while the rhythmical elements become more prominent in the mix. But the tonal elements are still playing a role, even though they do not sound all the time.

6.5 Prospects

A similar argument as proposed in Chapter 5 can be made here: All the points of critique do not render the analysis useless. And they should not make us dismiss the analysis in its entirety. When you look aside from the disturbing factor of the media's interpretations, and the authors' bombastic statements, the study can actually teach us something about the history of popular music. But for reasons that are not explicated in the text. These reasons are found in the design of the study, in statistics, and in the large amount of output.

6.5.1 The benefits of the approach

The data-driven approach

The study's high degree of being data-driven is a way to examine music without prior hypotheses. Like Rodriguez Zivic et al.'s (2013) analysis of tonal information in classical music, this analysis is designed as an unsupervised analysis. The researchers initially have setup the ways of retrieving the data, and the algorithms form both the topics and the styles from this data. Thereby the analysis largely eschews human interpretation, because it allows us to let the machine do the divisions of music into categories. After the algorithms have clustered the music, "human" words were ascribed to each cluster.

I am not arguing that these divisions thereby become more valuable than if humans did them but rather that the algorithms enable new views of our corpus. Perhaps the methods in combination with large datasets could provide us with information that could empirically confirm general theories, as in the case of Rodriguez Zivic (2013), in which the algorithms clustered the music in accordance with prevailing categorizations of musical periods. Or perhaps the algorithms could have lead to revealing hitherto unknown patterns. So yes we can use data to “go beyond what music experts tells us” as Mauch explains in his TED Talk about the study. But for the sake of music analysis, some expertise to both set up the algorithms and understand the patterns stand for is requisite.

From the specific to the general

As I argued in 6.3, the study is more specific than it accounts for in the text. But this specificity is what empirical research always has to do; examine small corners of the world. The corner analyzed here is not irrelevant to popular music research, because the relationship between the music in this study and popular music in general (whatever that is) is not arbitrary. As such, the US chart is a good indicator of an important part of music history and therefore not uninteresting. In fact, it is a larger corpus than most other studies, and thereby it provides a broader view than have been previously possible.

Statistics

The algorithms have distributed the majority of songs into categories that seem reasonable. Therefore it is most likely also acceptable if the algorithms calculate isolated mistakes in relation to human categorization, such as the examples I provided above. The study does what big data and statistics are good at; it creates compact descriptions of the world. And it provides a comprised and crude overview of certain aspects popular music history in the USA; of how things have sounded and overall trends and tendencies. It also points us towards interesting music historically events; that something interesting may have happened in the charts around 1991, 1983 and 1964. But musicologists may still argue that the major general changes happened in other years. And they may be right.

Statistics are well-suited for nuancing too; they can demonstrate that history is not as black and white as we might have a tendency to consider it. The developments in topics indicate that a certain sound idiom emerged in the charts in the early 90's, but this was not all that happened these years. Statistics enables us to get rough estimates of how large a share of the total number of songs that could be categorized into this sound idiom. Thereby statistics can help us nuance the main narratives of history by telling us that changes are not so subversive or overwhelming as they sometimes can seem to be in retrospect.

Visualisations

A further consequence of statistics is visualization. The graphs make these overall tendencies easy-to-read. As a non-expert in American chart music, they helped me get a better grasp of something, which I became more aware of after reading the study. But I could not let data work alone; I had to mix in knowledge about music analysis, statistic methods and popular music history in general.

Visualizations suggest a quick way of overviewing complexities. They can remind scholars and students that things have changed, despite changes happen slowly. And they can suggest what changed and when. To some extent, they have a similar function as imprecise maps. In a map, you only get details about certain aspects of the world, namely the geographical. Just like a country is way more than just a share of land, music is way more complex than can be fitted into eight harmonic and eight timbral topics. But you get a sense of overall directions and tendencies because in both cases a lot of other musical factors tend to follow indirectly. Songs that share timbre and tonal properties also tend to sound the same in other ways, rhythmically, phrasings, melody, etc. Correspondingly when looking at a map, countries closer to each other tend to have more in common than those far apart. If we look at Europe, there are, for example, differences compared with Africa, demographically and culturally. However, domain knowledge will also teach us that this is not always the case.

Reading the study as an exploratory one

I have already suggested that the study can be read as an exploratory one. This implies that the charts and the data can be read as suggestions rather than arguments. The data can then be a starting point for closer investigations. For example as the first step in an alternating process between human and machine interpretation: The findings can assist where to turn the attention, where to focus next and suggest a lot of new questions that I can pose. I can, for example, compare with other countries' popular music history. Or I can search for patterns regarding other musical parameters. Or manually listen closer into what happened in 1991. The study raises more questions than it answers. And that is not necessarily bad.

6.5.2 Benefits of the supplementary information

The research setup leads to the creation of a lot of data, which consequently also can be presented in databases, tables, and graphs. However, neither of the methodological steps

in the analysis are self-evident, and each step holds a potential of biasing the results in certain directions. Rieder and Röhle allege that “[i]n order for results to be challenged and critically assessed, there needs to be a high degree of transparency regarding assumptions, choices, tools, and so forth” (2012, 80-81). Rieder and Röhle suggest companion websites, which “seems like an ideal vehicle to present results more dynamically and interactively. Instead of providing the research results as a closed, finished product, Web-based interfaces could allow audiences to explore them both inductively and deductively, involving them in the process of knowledge production.” (81)

Mauch et al. do not provide a flexible, interactive companion website, which would enhance the understanding of the study lot, but they do provide a significant amount of supplementary information: The supporting information¹¹⁴ explains the methodology, but it requires a high level of understanding of MIR and mathematics to comprehend it. In-depth tables¹¹⁵ integrate the last.fm-data with the MIR features showing top artists and top genre tags for each of the 16 topics, and which chord progressions are grouped together in each H-topic. A full dataset¹¹⁶ of all songs’ topic values, PCA values, and harmonic and timbre values is accessible in a 17095x269 CSV file. And audio files¹¹⁷ containing sound excerpts of all timbre topics have been created.

The supplementary material helps to make the methodology more transparent. It allows musicologists to locate music they know and are capable of analyzing roughly from sheer reading the title and see how it is quantified and classified. The material thereby facilitates a better intuitive understanding of the large-scale calculations. For example are the top 5 T1 (drums, percussive, aggressive) artists TLC, Pet Shop Boys, Jody Watley, The Cars and Paula Abdul¹¹⁸, respectively. These artists illustrate well what timbre characteristics that comprise a common denominator within this topic: Despite being stylistically diverse, these artists represent a staccato sound with rhythmic and percussive elements salient in the mix. While hip hop, which I at first associated with the “drums, percussive, aggressive” label, typically is distributed into T3 (energetic, speech, bright) (but also T1). This knowledge takes us closer to an explanation of the graphs in Figure 19, presented above, which show the trends regarding each timbre topic.

114 <http://rsos.royalsocietypublishing.org/content/royopensci/suppl/2015/05/06/rsos.150081.DC1/rsos150081suppl.pdf>

115 <http://www.eecs.qmul.ac.uk/~matthiasm/descentofpop/fullTables.pdf>, retrieved October 1, 2016

116 https://figshare.com/articles/Main_Dataset_for_Evolution_of_Popular_Music_USA_1960_2010_/1309953

117 <https://soundcloud.com/descent-of-pop>, retrieved September 28, 2016

118 This playlist: <https://open.spotify.com/user/hviderusland/playlist/1v4tqZvBcTCOdy7Ay4HLKM>, contains songs with high T1: Number 1,2,4 and 5 on the list. The playlist contains the song with highest T1 value from each of the mentioned artists. (Except Jody Watley, whose second highest T1 scoring song is on the playlist)

Even though they are not completely reliable because of the calculations being based on excerpts, especially the in-depth tables and the dataset can serve as a heuristic catalog. They provide us with coarse estimations of the most typical chord changes in different genres and within artists' oeuvre. They could, for example, become a point of reference for musicologists or students not trained in identifying chords from sheer listening. Or they could play the role of a rudimentary starting point for other music analyses. Again this kind of information is most valuable when put in context. If for example analyzing an imagined new trend in pop music which applies a lot of major 7th chords, the in-depth tables tell us that these chords are typical in blues, jazz, funk, easy listening and the 60's, and less common in soft rock and hip-hop. This knowledge could be a part of an argument that claimed this new trend was inspired by blues in the 60's. And yes, we were also able to do this before digital techniques, but now this information is very easily accessible. The tables provide us with an easy way to get empirical information that can inform other types of analyses.

6.6 Conclusion

In this chapter, I have examined the epistemological value of a music big data study. I have chosen to investigate this study because it exemplifies well many general issues which often are at stake when conducting large-scale music analysis by means of digital techniques: The schism between research cultures, black boxing, and how to translate from data analysis to music analytic value. The aim was to scrutinize the applied methods to determine their potentials for future musicological research. Doing so, gave me reasons for questioning the study's epistemological value: Complicated methods complicate the findings. And many musicologists will probably dismiss the study because the statements presented in the text appear too bombastic and conclusive. The text tends to conclude more than the numbers indicate.

Paradoxically, even though this study holds a problematic epistemological value, some may still criticize it for not teaching us anything we did not know beforehand. That it is no big surprise that the sound changed in the Billboard Hot 100 around 1964, 1983, 1991, or that the amount of major 7th chords has decreased, for example. If that proves to be the case for other studies as well, I would be very optimistic about the tools, because it would indicate that they work as intended.

These methods would then work sufficiently well to become an integral part of a strategy on how to deal with the abundance of music now available digitally. Simply because they can quickly present knowledge that otherwise would take a lot of listening to retrieve. We will be able to retrieve a lot of fairly reliable information, and thereby get an easier grasp of music that we did not know already. For example, to retrieve crude statistics about music in various South American countries, and present them in visualizations easy to overview tendencies: From grasping the totalities to zooming in and exploring nuances of subsets of a country, comparing it with other countries, etc.

But epistemologically, we are still in a phase where we need more examples. The more we investigate, get hands-on experience, scrutinize in detail, we will become able to understand the caveats inherent in the particular methods used. We still need knowledge about what we analyze to understand what the calculations show us, and currently, we are limited by the fact that we are not yet accustomed to dealing with these methods. This preliminary stage of using MIR tools implies that they currently have questionable interpretational value. In Chapter 5, I argued that methodological considerations have to go into the foreground until best practices have been established, and this appeal was only enhanced by investigating Mauch et al.'s analysis.

The study also demonstrates that the danger of “visions of objectivity and universalism that hitherto had little currency outside of certain fields might gain momentum via an unreflected enthusiasm for technology” (Rieder and Röhle 2012, 80) becomes tangible and needs a counter-response. Another concern is, as Rieder and Röhle also warn, that “visual rhetoric and technological black-boxing form a conglomerate that is difficult to disentangle and has far-reaching epistemological consequences” (80). The study is an example of a conglomerate of technological enthusiasm and black boxing, and the consequences of handling the music the way Mauch et al. do is to a large extent incomprehensible.

I, therefore, request that we are honest about the methods' limitations. They suggest us modes of grasping corpora, as a strategy for dealing with the abundance of sources that was not possible pre-digitally. But if we are too concluding and non-critical about what they show, we risk distorting and impeding a dialogue about the methods proficiency.

CHAPTER 7

A Corpus Study of 89 DJ Sets

As described in the previous chapters, there are not many good examples of digital musicological research applying ACA methods for investigating large amounts of music. There is also a general lack of practical, empirical knowledge about to what extent the methods can be of any assistance in musicological practices. The primary purpose of this chapter is therefore to practice digital musicology with ACA methods. Through practice, I will investigate how well ACA methods are for informing music analytical relevant questions. These questions concern professional DJs' music selection practices; what music do they play and in what order?

For that purpose, I will deploy some of the most commonly used ACA methods to examine whether they hold music analytic potential. Next, I will analyze whether the big amounts of data the ACA methods creates can answer questions that concern the music. In short, what new potentials arise, when we apply ACA methods to pose and answer music analytical questions? And what qualitative answers do the measurements¹¹⁹ entail? As I outlined especially in Chapter 3, the focus is on translating from the quantitative measurements to musical qualities, since I pursue bridging the semantic gap between the computer's machine reading and human interpretation.

The chapter will be formulated as a step-by-step demonstration that has the aim of exposing and elucidating how and interplay between data and music analysis can be attained in practiced. How data can assist the exploration of the corpus. How the features can be converted to music analysis. And how data analytical techniques can be applied to both generalize and nuance.

¹¹⁹ As I explained in Section 1.7, the words *measure*, *measurement*, and *metric* are applied in their statistical meaning.

7.1 Introduction

7.1.1 Choice of case

For my analysis, I chose 89 DJ sets played and recorded at the Electronic Dance Music (EDM) festival Ultra Miami Festival (UMF) in March 2015. This Festival staged a significant share of the world's top-earning DJs. According to Forbes' list¹²⁰ from 2015 of the highest paid DJs, 8 out of top 10 played at UMF. This festival is one of the world's biggest annual music festivals with a total attendance of around 165.000¹²¹.

DJs' music selection was chosen for this case study because I presumed that this music genre was ACA-friendly, and therefore it would enlarge the chance of success: Electronic Dance music (EDM) has a supposedly acoustical measurable way of creating musical development. Most EDM is composed of repetitive structures¹²², and musical progress is often created on levels that can be identified from looking at a spectrogram. One way of creating musical progress is, for example, to introduce a new instrument into the loop, most prominently by introducing or re-introducing the bass drum¹²³, which affects the spectrogram and the envelope profoundly.



Figure 22 Poster from UMF 2015: DJs have their own logos and appear on this festival poster, just like other artists would do on any music festival poster.

¹²⁰ <http://www.forbes.com/electronic-cash-kings/> This list includes all earnings, and not just those earned by DJ'ing.

¹²¹ <http://www.emusician.com/artists/1333/ultra-music-festival-announces-2015-aftermovie-2016-ticket-options/54296> , retrieved May 1, 2017

¹²² See for example (Björnberg 1996) for an analysis of DJ Seduction's "Hardcore Heaven".

¹²³ This is both a very common and very effective maneuver: As Butler (2006) explains: "The most common phenomenon involving the removal of the bass drum - followed, of course, by its eventual return. This dynamic of removal and return is pervasive within EDM, appearing at some point in nearly every track." (91). For anecdotal indication of the effect of this phenomenon, read for instance the introductory scene Butler, 2006, in which Butler exemplifies the power of DJ Stacey Pullen's cutting out the bass drum, and bringing it back on, and its impact on the audience (3).

It was important for me that the corpus was manageable in terms of size: I wanted to be able to listen to all sets and to shuttle between data analysis and music listening back and forth. At the same time, I wanted the corpus to be so big that my mental capacity would not be able to handle all details and nuances, which consequently would render computer techniques appropriate for the task. At UMF 2015, a total of 170 DJs and live acts¹²⁴ were presented on seven different stages, each with a different focus representing styles within EDM. I was able to retrieve audio files containing music from 89 of the acts¹²⁵. These represent 52% of the artists who performed at the festival. Main Stage, Ultra Stage, Live stage and Resistance Stage are stronger represented than the other stages. Recording methods and audio quality vary, which is a factor that plays a role in the analyses. When more recordings were available first priority was length, second priority was the quality of the recording. Each audio file was compressed as a wav-file and split into segments of 10 minutes. Some of the analyzed audio files contain sound that does not adhere to the performance itself, most noticeable speech from artist interviews.

7.1.2 Related research and theoretical bases

Brewster and Broughton have boiled the craftsmanship of DJ'ing down to its core: "The essence of the DJ's craft is selecting which records to play and in what order" (2000, 9). This essence is also my point of departure in this analysis. I will apply ACA methods to investigate *what music* the DJs play, and in *what order*. Fikentscher has remarked that "there is a strong rationale to examine DJ programming [...] at the microlevel," while stating that a predominantly "macro-level approach, focusing on broader themes and relationships" have been foregrounded in the academic discourse regarding DJs (2013, 124). These macro-level approaches encompass relating DJs and DJ'ing to more general issues such as gender, identity, technology, the relationship between human and technology, postmodernism, genres, investigation of specific scenes, etc.

Langlois (1992) and Hadley (1993) have written early introductory texts to the dance music scene. While other perspectives on DJ'ing amongst others include DJ history (Brewster, & Broughton 2000; Brewster, & Broughton 2010), authorship (Herman 2006), how to DJ-guides (Broughton, & Brewster 2003; Brophy, & Frempong 2010). In-depth analyses of EDM tracks are found in (Björnberg 1997; Butler 2006). The former focusing on tonal modality, the latter conducting in-depth rhythm analysis of various EDM subgenres.

¹²⁴ The distinction between DJs and live acts is blurry, because DJs often play their own music.

¹²⁵ The editorial processes for the criteria for which mixes have been recorded and made online is unknown to me.

Regarding the craftsmanship of DJ'ing, there seems to be a common agreement on what musical skills it takes to be a good DJ¹²⁶. A good DJ should have a well-developed awareness of energy, emotions, mood, and atmosphere of both music, the audience and the room and how these relate and affect each other (Broughton, & Brewster 2003; Broughton, & Brewster 2003; Fikentscher 2013; Langlois 1992). The DJ's awareness of these aspects manifests itself in the music he or she plays. These parameters, however vague and difficult to measure, are in the background of my analysis. Each DJ has his own set of ideals on which energy and which atmosphere, and how to create them. I will seek to uncover some of these tactics and ideals, and how they manifest in the music.

Fikentscher has conducted an ethnographical study of music selection (2013). And Montano (2009) has interviewed DJs in Sydney on various topics, some of these regard music selection practices. Greasley & Prior have investigated DJs' relation to shape on both micro- and macro-level (2013). I will examine different approaches to shaping whole sets in Section 7.4.

Kell & Tzanetakis (2013) have empirically analyzed track selection choices, and track order in EDM mixes played in BBC radio. They applied ACA methods for analyzing to what extent a succeeding track holds the same tempo, tonality, and loudness as the previous one. They found a stronger cohesion on these parameters between two consecutive tracks played by a DJ than when the same tracks are played in random order. In comparison to their analysis, the goal of this analysis is also to investigate and identify music selection strategies. However, I am seeking to compare and discern them from each other, to be able to determine stylistic differences between for example DJ Hardwell and Tiësto, corresponding to being able to distinguish Haydn from Mozart.

7.1.3 Informing MIR

Like Kell, & Tzanetakis explicitly state, I too hope that my study, as a possible side effect, can inform the generation of MIR recommendation algorithm with musicological insight. Music recommendation is a MIR key task (Celma 2010) and the DJs I investigate have through years of practice developed a strong intuition for which music to choose, and how the choices affect an audience. In this analysis, the music will be analyzed by "looking at it" through some of the same metrics as music recommendation algorithms do when they

¹²⁶ While it also take other than purely musical skills to become a star DJ, such as personal branding (Sherburne 2012) and performative qualities). Rietveld and Reynolds have argued that visual aspects of especially superstar DJs performances have gained increasing prominence and importance (Rietveld 2013; Reynolds 2012). Nevertheless, the music is still to be considered the core product (Fikentscher 2000; Small 1998).

automatically recommend the next track. However, at the same time, these DJ sets I investigate comprise an extreme case of music selection: The purpose of the music is to make people dance, it is a concert-like setting, and the music has high attention from a very participating audience. Besides, the DJ sets involve a lot of manipulating of the tracks played. These creative techniques applied by the DJs are to my best of knowledge not yet reproducible by current technologies.

7.1.4 Choice of Methods

In his essay *Postproduction* in which the DJ is one of the central figures covered, Bourriaud wrote about forming artworks out of already existing ones that “[t]he material they manipulate is no longer *primary*” (2005, introduction - emphasis in original) Theory more specific concerned with DJ’ing suggests similarly. Butler, for example, states that “[a] set is a unity [...] the emphasis is on the larger whole rather than its components” (2006, 49). While Fikentscher understands music programming to include “the strategic control over tempo, pacing, selection of repertoire and sound effects, including even the manipulation (emphasis or de-emphasis) of frequency bands” (2013, 125). In this analysis, I will accordingly analyze each DJ set with all its “strategic controlled” components. Therefore, I regard as a whole, one *composition*¹²⁷, and not as its separate tracks. The objects of analysis, therefore, become the recorded audio files of the performances. This corresponds to what the audience experience: If a track played has a fall in energy, the final output will have the same fall in energy. Unless, of course, the DJ manipulates it to prevent this fall.

For feature generation, I chose MATLAB based software MIRtoolbox (Lartillot et al. 2008). It met all the requirements I had before the analysis, since it can retrieve a large variety of the most commonly applied ACA features from audio files. The collection of features in MIRtoolbox includes *tonal*, *rhythmical* and *timbral* features, and both traditional music analytic measures and born-digital measures. Like Echo Nest, MIRtoolbox enables the creation of a feature vector space for each audio file. This allows me to approach the music from a data analytic approach, in which many musical aspects can be taken into consideration and investigated using data analysis and visualization. But in comparison with Echo Nest, MIRtoolbox can create more features. Many of these are low-level which implies a more transparent relation between measurements and music. MIRtoolbox,

¹²⁷ The word *composition* has multiple connotations attached to it. In Oxford Music Online composition “describes a process of construction, a creative putting together, a working out and carrying through of an initial conception or inspiration”. In this thesis, the word is applied in a more generic sense, as Butler (2006, 50) suggests. He refers to the Latin root *componere*, “to put together.” Correspondingly the Oxford Dictionary defines *Composition* as “A work of art [...], consisting of several elements artistically combined.” (21c). It is in this sense I apply the word.

therefore, allows me to investigate the more basic building blocks of ACA, and how these measure musical aspects. Furthermore, MIRtoolbox provides many measurements of each audio file each second, which enables breaking the music up into subtle passages of the desired length, ranging from millisecond-level to the duration of the entire file. The customizability and flexibility are a lot higher with MIRtoolbox, but it also requires a lot more technical and mathematical expertise to engage with the software.

7.2 Step 1 - Exploring the Features

7.2.1 Point of departure: A lot of data

With default settings, MIRtoolbox calculates feature values for every minuscule time interval, several measurements for each second. For this analysis, this led to the creation of many large datasets for each audio file containing feature values: One file for each feature for each 10-minute segment.¹²⁸ This very large amount of data was very hard to manage in its raw form. I discarded features that were very difficult to handle, often due to too high-dimensional output, and those that lacked music analytical relevance in relation to the questions I had. To be able to retrieve music analytic insight from the data, I had to investigate each of the features to understand what they represent. For a few of them, I found it useful to modify and calculate new features that were more meaningful in relation to the music I was investigating, and the questions I had.

For some features, the feature value corresponds with certain aspects of how the music sounds. For these features, the link between acoustic aspects of the audio file and feature values is rather understandable. For other features, this link between description and measurement is less obvious. But this does not necessarily render these features useless, and in some cases, I chose to include them in the analysis despite dubious explanatory value. One reason is that more data is easy to manage digitally and may not be problematic. The PCA, for example (see Section 5.6.2), allows me to investigate whether a given feature is significant for structuring my dataset. MFCC's are examples of such features that have no palpable connection to the experienced sound, but nevertheless have proven useful for various MIR tasks.¹²⁹

¹²⁸ The total amount of features created for the DJ sets piled up to 21GB in CSV form. They can be found here: <https://zenodo.org/record/806207>

¹²⁹ However, for this analysis, I tried to include MFCC's, but did not find them fruitful in relation my purpose.

The first data analytic step in this analysis was to examine the relationship between features and music. My primary question was *what musical aspects does a given feature measure?* I investigated this by comparing the MIRtoolbox manual's description of the features with feature values of a reference corpus, consisting of 18 songs within a wide range of music. The reference corpus consisted of:¹³⁰

2 Unlimited – No Limit

Major Lazer (feat. MØ & DJ Snake) – Lean On

Portishead – Glory Box

Adele – Hello

Mark Ronson, Bruno Mars – Uptown Funk

Radiohead – Everything is in the Right Place

Bob Marley & The Wailers – Could You Be Loved

Michael_Jackson – Billie Jean

Igor Stravinsky – Firebird VII: Game of the Princesses with the Golden Apples

Daft Punk – Around The World

Miles_Davis – So What

System of a Down – Chop Suey

Ipelegeng Experimental Group – Meropa Le Di Kota

Nirvana – Smells Like Teen Spirit

White Noise – #1 minute of white noise#

Jay Z & Kanye West – Niggaz in Paris

Ockeghem – Requiem: Kyrie

Madonna – Like A Prayer

Pharrell Williams – Happy

These songs are biased towards EDM's stylistic characteristics. There are more popular music than classical music, and there is more dance music than ballads, more electronic than rock, and more recent music than old. I chose to investigate the metrics against a more general corpus with a broader range of music in order to widen the perspective on how the features "behave." Expanding the range of genres would enhance the number of caveats that I could reveal. As an extra bonus, this approach also offers insight into about what features EDM music hold when compared to other genres. I will eventually also refer to a DJ test corpus, which comprises of 10 DJ sets from minute 10-20, from various EDM subgenres.

¹³⁰ Spotify playlist: <https://open.spotify.com/user/hviderusland/playlist/0CuXAzHqCHuLtKUFZNHUttr>

For a few of the standard features, I found that it could improve my analysis if I modified them. Consequently, I created a couple of new metrics for my analysis. The way I approached this task exemplifies well how ACA “listens” to music and how to exploit this to create new analytically meaningful metrics. It also demonstrates how one’s music analytic mindset has to be adjusted to conform to the levels of milliseconds, and how one can apply mathematical music analytical thinking to create music analytic inquiries. Therefore I will explain the process of creating these.

7.2.2 Rhythmical features

The rhythmical features comprise of *Lowenergy*, *Eventdensity*, *Tempo*, *Metroid*, *Beatspectrum*, *Pulseclarity*. Only *Tempo* and *Eventdensity* are detectable without digital methods, with tempo by far the easiest intuitively perceivable. To MIRtoolbox, *rhythmical* implies measurements of aspects of the sound signal’s energy curve. These features are thereby largely ignoring music’s tonal and timbral aspects.¹³¹ This means that the concept of rhythm is not only limited to drum sounds but takes the whole acoustic image into account. In comparison with the analysis of scores, the length of a bass drum will in this type of analysis influence the features, simply because the energy curve is dependent on the length of the sounds. None of the features I apply measure rhythmical aspects of time units longer than a second. This analysis part is therefore very oriented towards rhythmical aspects of the individual beat and not about how these individual beats form groups. Metrical aspects (in the music analytic meaning of the word) are omitted, which is not due to MIRtoolbox capability of providing measures for it, but because I found the feature output too difficult to handle and make useful for my purpose.

Lowenergy

A thorough presentation of how I approached and altered MIRtoolbox’ *Lowenergy* feature exemplifies well the type of thinking that has to be applied when translating from measures to music analytical value, and hereafter creating new more music analytic meaningful metrics. *Lowenergy* is a measure of how many frames that show less-than-average energy measured as root-mean-square (RMS) energy (Lartillot, 2014; Tzanetakis, & Cook, 2002). *Lowenergy* is calculated from RMS values for each frame of 50 ms. In a song that holds a tempo of 120 BPM, there is 500 ms between each beat, i.e. ten frames per beat. Therefore, the sole sound of a bass drum, lasting less than 50 ms, only playing at each quarter note at tempo 120BPM will imply a *Lowenergy* value far above 50%, because the majority of frames has less than average energy.

¹³¹ Though tempo also rely on spectrum analysis.

Investigating the reference corpus helps me identify that some concerns will arise when attempting to translate from Lowenergy values to musical aspects. The problem for creating a transparent relation between feature and music is that MIRtoolbox in the default settings provides only *one* Lowenergy value for each entire audio file.

Statistically, extreme values affect the mean much more than they affect the median. In relation to the Lowenergy value, the difference between mean RMS and the median RMS must, therefore, correlate with Lowenergy, because if the mean is larger than the median, there are more frames that are beyond mean than above. Think again of the example of the sole bass drum sound lasting one frame, in this case, the median over 1 second will be 0, but the mean above 0. But since the majority of frames have 0 energy the Lowenergy value will be almost 1.

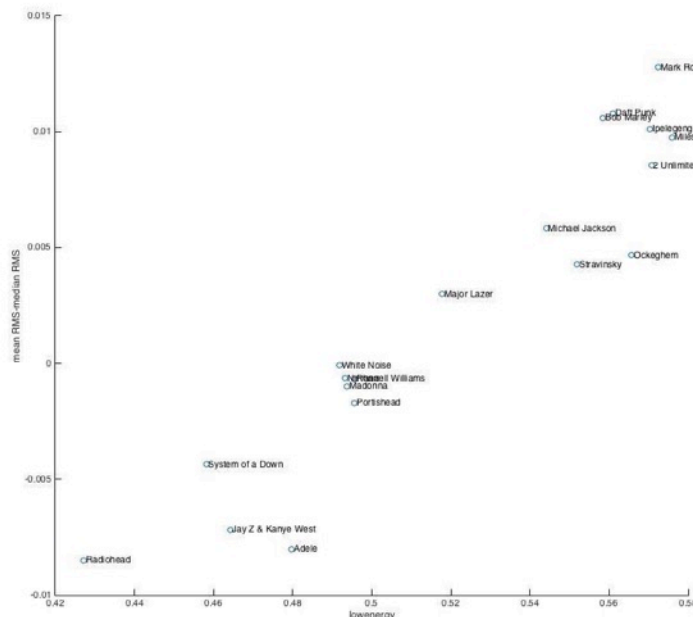


Figure 23 Correlation between mean RMS-median RMS and lowenergy values for the reference corpus.

Looking at the RMS energy curve of Ockeghem provides us with some information on its high Lowenergy rate:

The peaks are formed by dynamic fluctuations and are very high compared to the rest of the signal. Statistically, they affect the mean more than they affect the median and therefore the mean RMS is larger than the median RMS, resulting in more frames below mean, and thus a Lowenergy value above 0.5.

Miles Davis' *So What* is another case, in which the overall dynamic development of the song plays an important role for the Lowenergy value. Miles Davis has the highest Lowenergy value in the reference corpus.

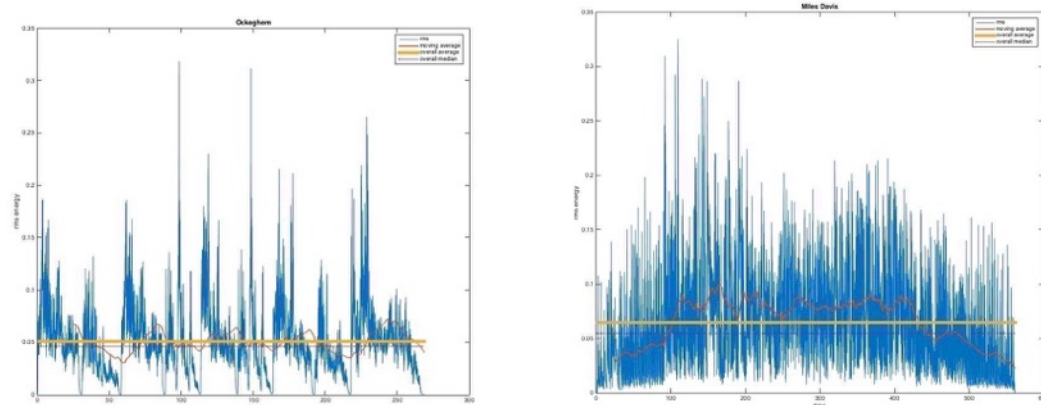


Figure 24 The RMS energy curve of Ockeghem and Figure 25 The RMS energy curve of Miles Davis. The yellow line displays the average. The red is moving average, the dotted blue is the median.

Local Lowenergy (LLE)

Ockeghem is an extreme case, compared to the corpus I am investigating, because it is dynamically far from the idiom of EDM. These concerns about dynamic peaks influencing the results are presumed much less apparent in EDM where the audio signal typically is very compressed, implying less dynamic changes. However, Figure 26 exposes that there are dynamic variations within EDM. The graphs display RMS energy curves of the first 10 minutes of a subset of the DJ sets. Especially the plots of *Boys Noize* and Adam Beyer - *Ida Engberg* unveil that there are both energy fluctuations at millisecond level and at a larger scale that would render the musical analytical value of Lowenergy equivocal.

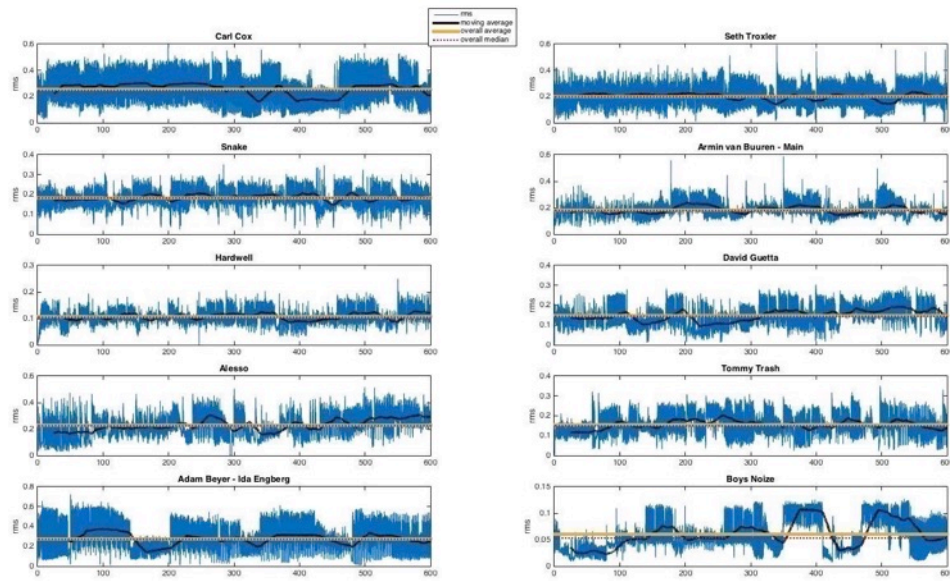


Figure 26 Plots of the energy curves of the first 10 minutes of a subset of the DJ sets.

In order to enhance the musical analytical value of Lowenergy, I created a new feature, which focusses sheerly on calculating Lowenergy values for every smaller time units. I created the feature to ensure that it solely measure musical characteristics that adhere to the beat level and eschew large-scale dynamic issues.

Local Lowenergy (LLE) measures how many frames in each second that shows less-than-average RMS energy for this second.

LLE¹³² is an indicator of the amount of dynamic contrast at the one-second level. The *low-high* contrasting dynamic in Nirvana becomes apparent in LLE, as demonstrated in Figure 27: In the verse the sound is more sparse containing a loud snare drum, resulting in more dynamic variation at one-second level. The chorus contains denser sound, drawn out distorted guitar sounds, and less dynamic fluctuations, and therefore also lower LLE values.

Stylistically, it is remarkable is that the R^2 ¹³³ value, expressing the degree of correlation between MIRtoolbox' original Lowenergy value correlates and the new LLE feature, is

¹³² Script: getLLEdata.m

¹³³ The higher the R^2 value the more correlation. R^2 is always between 0 and 1.

0.237 for the reference corpus while it is 0.838 for all DJ sets in the datasets.¹³⁴ This indicates, not surprisingly, that these dynamic issues over the course of longer time units, I have explained above, are not as apparent in the DJ sets.

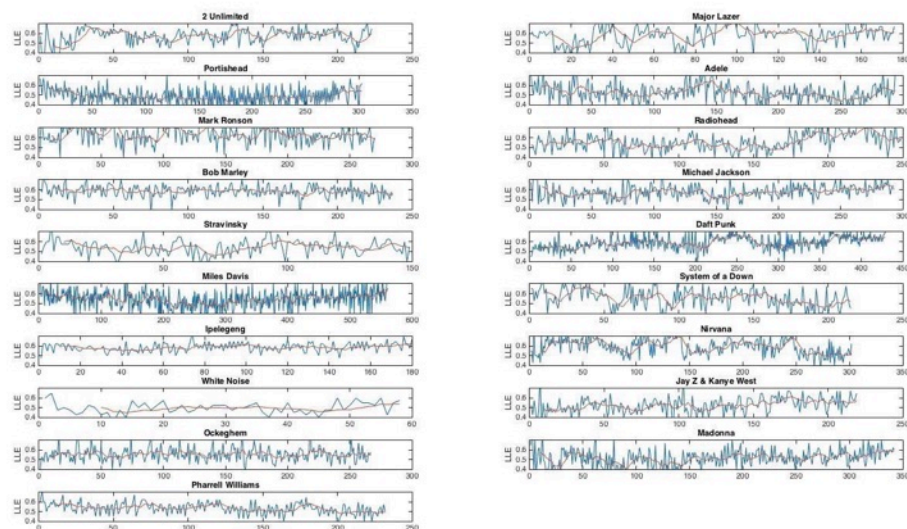


Figure 27 LLE values for the reference corpus. The red line is the moving average for 10 seconds intervals, thus there is a 10 seconds lack in this line.

Figure 28 shows a plot of all DJ sets in the dataset, from minute 10-20. The colors represent the stage of the act. The plot indicates that artists that played at the Resistance stage all played music with high LLE values. This indicates that LLE values to some extent can discern at least some EDM genres from others.

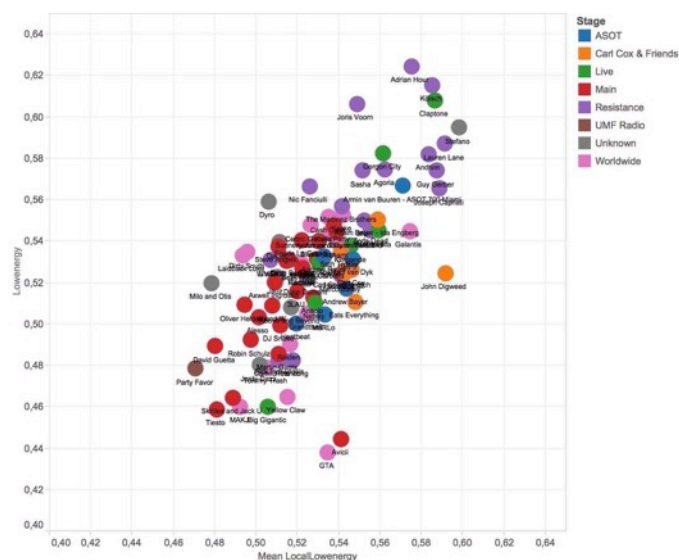


Figure 28 Mean LLE and Lowenergy for the DJsets minute 10-20.

¹³⁴ See Appendix I for a plot of the reference corpus and 10 DJ subsets.

Tempo¹³⁵

Figure 29 displays mirtempo¹³⁶ values for the reference corpus. Combining the graphs with knowledge about the music it becomes apparent that the mirtempo graphs indicate more tempo changes than the music suggests. The general problem for ACA methods is to identify which onsets mark where one would tap the beat, and which mark subdivisions or syncopations. "Simulating this cognitive process with an automated beat tracking system is much harder than one may think" as Müller formulates it (2015, 303).

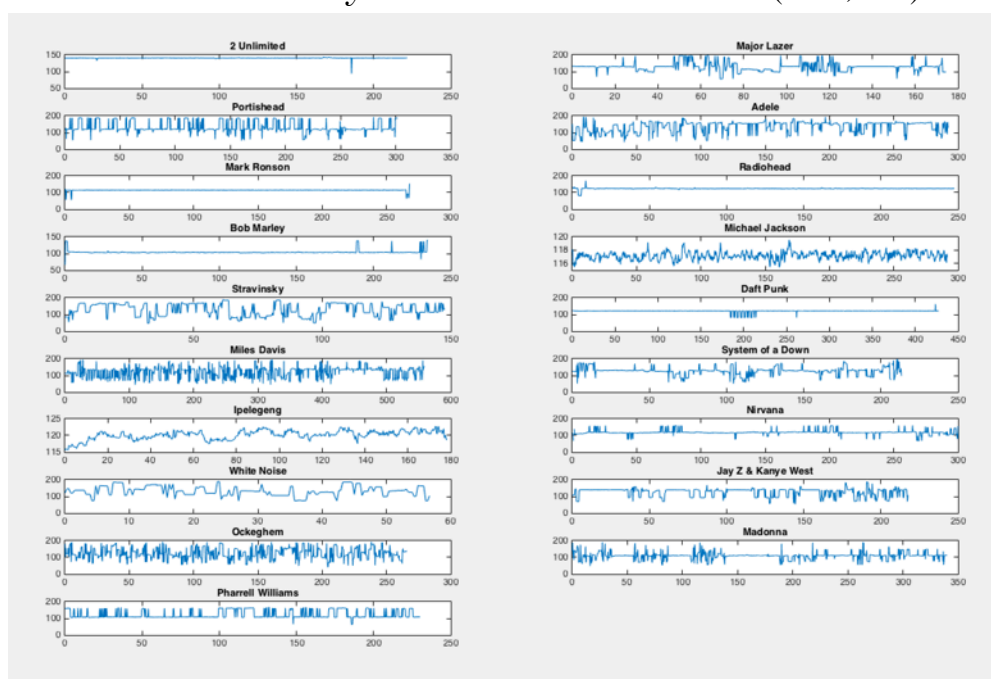


Figure 29 The reference corpus' mirtempo values calculated as a function of time.

But can this information be modified into something more music analytically meaningful? And are there other kinds of useful information for the task at hand, despite the mismatch between the tapping tempo and Tempo values?

For the task at hand, it is good news that the tracks with four-on-the-floor¹³⁷ (Daft Punk and 2 Unlimited) are tracks with a good match between humanly perceived tempo and mirtempo calculations. This is good because the majority of the music in the DJ set corpus also have four-on-the-floor rhythms. Daft Punk's and 2 Unlimited's tempo graphs are both fairly straight indicating only a little insecurity. Mark Ronson, Bob Marley, Michael

¹³⁵ Script: getTempoData.m

¹³⁶ MIRtoolbox' calculations of tempo.

¹³⁷ A bass drum hit for each quarter note, four times per measure. Concept explained in (Butler 2006, 78)

Jackson and Ipelegeng also have very straight lines¹³⁸, which also indicates a relatively stable estimated tempo. These songs hold a steady tempo, sounds with percussive qualities are predominating, and syncopations¹³⁹ are not prominent.

However, 2 Unlimited's temporary decrease at minute 3.06 is an example of a local deviation from the prevailing tempo. It is caused by a few 10th of a second, containing pure rap and a hi-hat. The machine probably "hears" the rap above the hi-hat and estimates the tempo according to the rap, despite that the hi-hat maintains the pulse. A similar situation for Daft Punk, after 3.00, where the voice divides two measures into 3 and five beats, respectively. This makes the system become in doubt whether the tempo is 120 or 3/4, i.e. 90.

Calculating average tempo makes dubious music analytic sense. The perceived tempo would more likely be either one of the values calculated by MIRtoolbox, and not somewhere in between. As a simple solution to this problem, I chose to calculate the *mode*¹⁴⁰ of the *mirtempo*. I found this number to be a more appropriate metric better connected to the perceived tempo. The mode corresponds to the highest peak in each histogram in Figure 30. In the dataset, I analyze, *MeanModeTempo* denotes the mean of each minute's mode of the tempo.

On Syncopation

Returning to the tempo graphs in Figure 29, I find it relevant it is also interesting to compare MIRtoolbox' calculation with how I would tap the beat to the song. For example, MIRtoolbox mostly estimates the tempo of Pharrell Williams as 108, which is 2/3 of how the snare drum, playing on 2 and 4 in each bar, indicates the beat. Major Lazer is on the contrary estimated 4/3 faster than the tempo of its bass drum, which plays a regular four-on-the-floor rhythm, indicating the tempo this way. Both (mis)calculations are most likely due to a lot of syncopated elements prominent in the mix. For example, Major Lazer's tempo estimation is most likely caused by a lot of dotted eighth notes in the played by the synth.¹⁴¹

¹³⁸ The visualization may trick you, because the Y-axes differ. If the Y-axis was identical to the other plots these songs' tempo curves would also be perceived as a straight line.

¹³⁹ Syncopation is in this chapter used as a general term meaning "a disturbance or interruption of the regular flow of rhythm" (Hoffman 2005, 239).

¹⁴⁰ *Mode* is applied here in the statistical meaning of the word: The value that occurs most in the data.

¹⁴¹ Further more, Adele and Portishead are mostly estimated at doubled the perceived tempo, thus exemplifying the recurring double tempo issue, I also covered in Chapter 5. While both Jay Z & Kanye West and Radiohead can be perceived as either 140 and 124 respectively, or the half; in both cases MIRtoolbox chooses the double tempo.

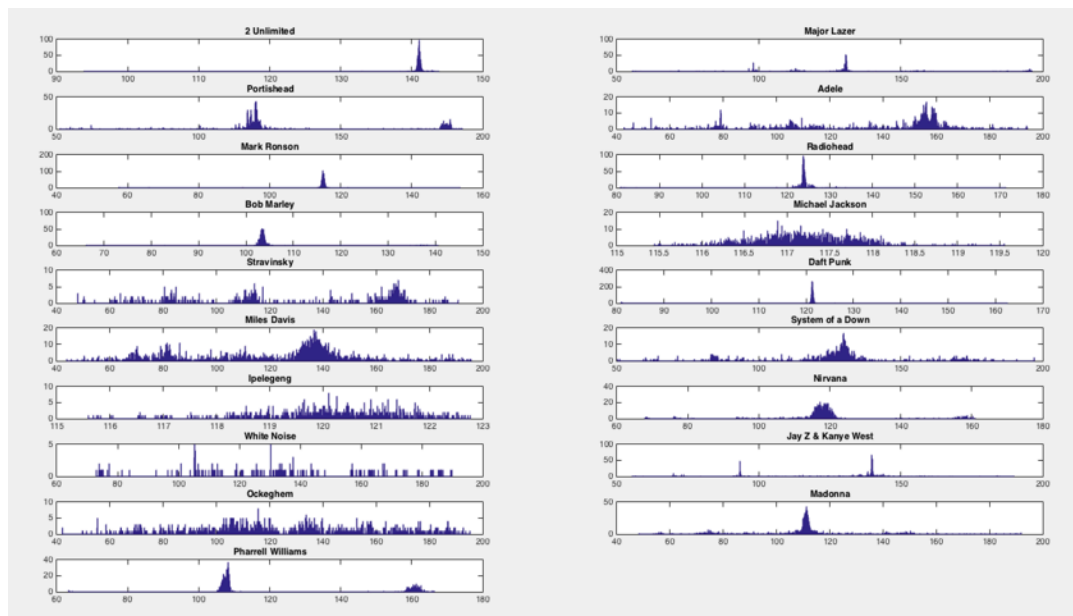


Figure 30 Tempo histograms of the reference corpus. Both X and Y axes vary.

Therefore, just as in the case of Echo Nest (see Section 5.4.5), I found indications that the computer's insecurity about the tempo held valuable rhythmical information. Therefore, I set up a new metric, which I call *TempoSynco*.¹⁴²

TempoSynco denotes the percentage of tempo calculations that within each minute is found within this minutes mode of the tempo ($\pm 12,5\%$), or within the doubles or the halves of the minute's mode¹ ($\pm 12,5\%$). TempoSynco is a coarse indicator of how much a beat is based on regular subdivisions, such as half notes, quarter notes, eight notes, etc. Songs with accentuations on these regular subdivisions score highest, while songs with accentuations outside these, such as on dotted notes score lowest. Songs with low TempoSynco tend to have more syncopated beats or that lack clear pulsation.

¹⁴² line 119-158 in `getTempoData.m`

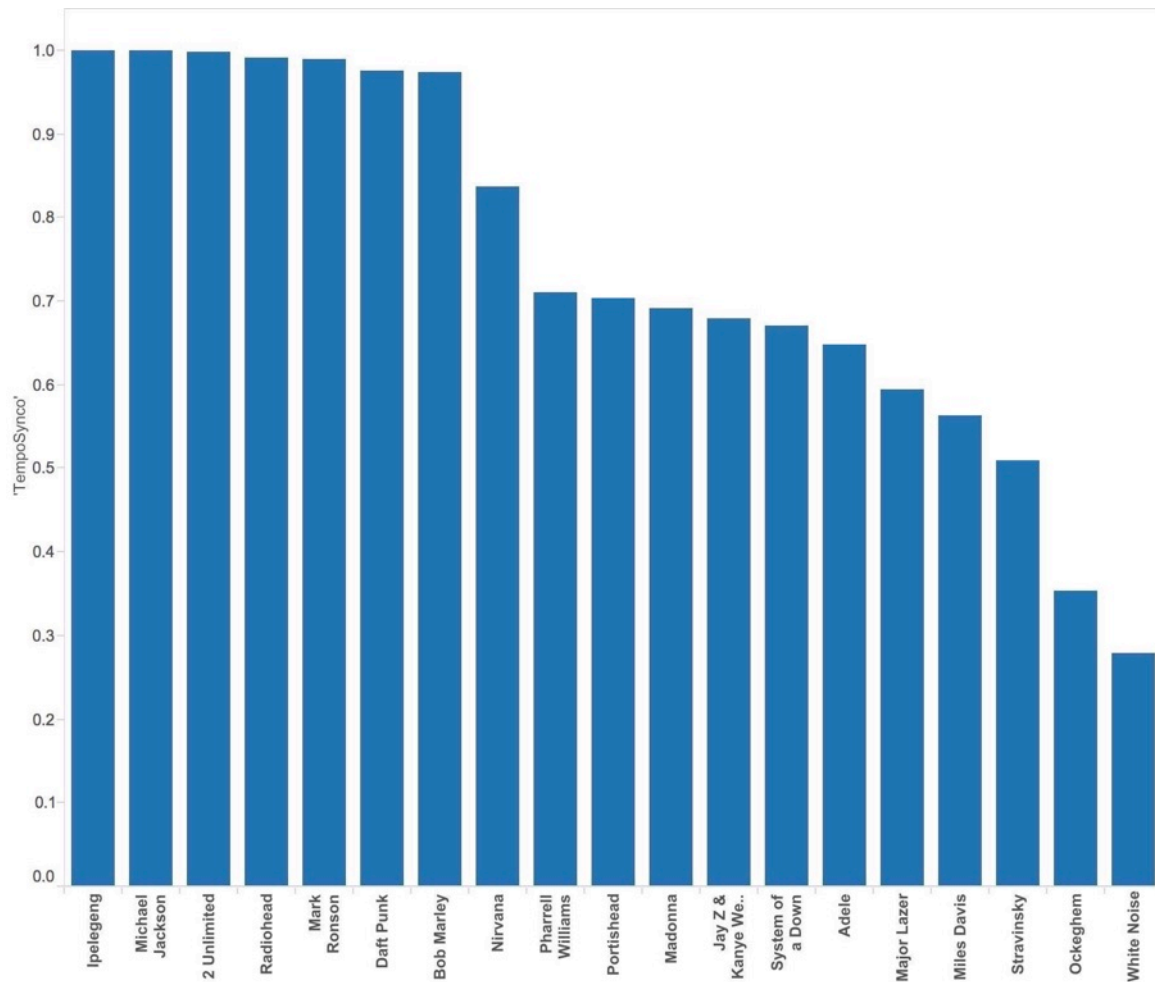


Figure 31 TempoSynco measures of the reference corpus.

Metroid¹⁴³

Metroid is a tempo related feature that indicate whether fast or slow metrical levels predominate:

“High BPM values for the metrical centroid indicate that more elementary metrical levels (i.e., very fast levels corresponding to very fast rhythmical values) predominate. Low BPM values indicate on the contrary that higher metrical levels (i.e., slow pulsations corresponding to whole notes, bars, etc.) predominate” (Lartillot 2014, 149).

¹⁴³ getMetroiddata.m

Figure 32 indicates that there is a connection between songs with rhythmical subdivisions prominent in the mix and high Metroid values. Both Daft Punk's and Bob Marley's metroid values are twice their tempo values, in accordance with their accentuated offbeats. In contrast, Radiohead, Portishead and Adele have low Metroid values corresponding to fewer accentuated subdivisions.

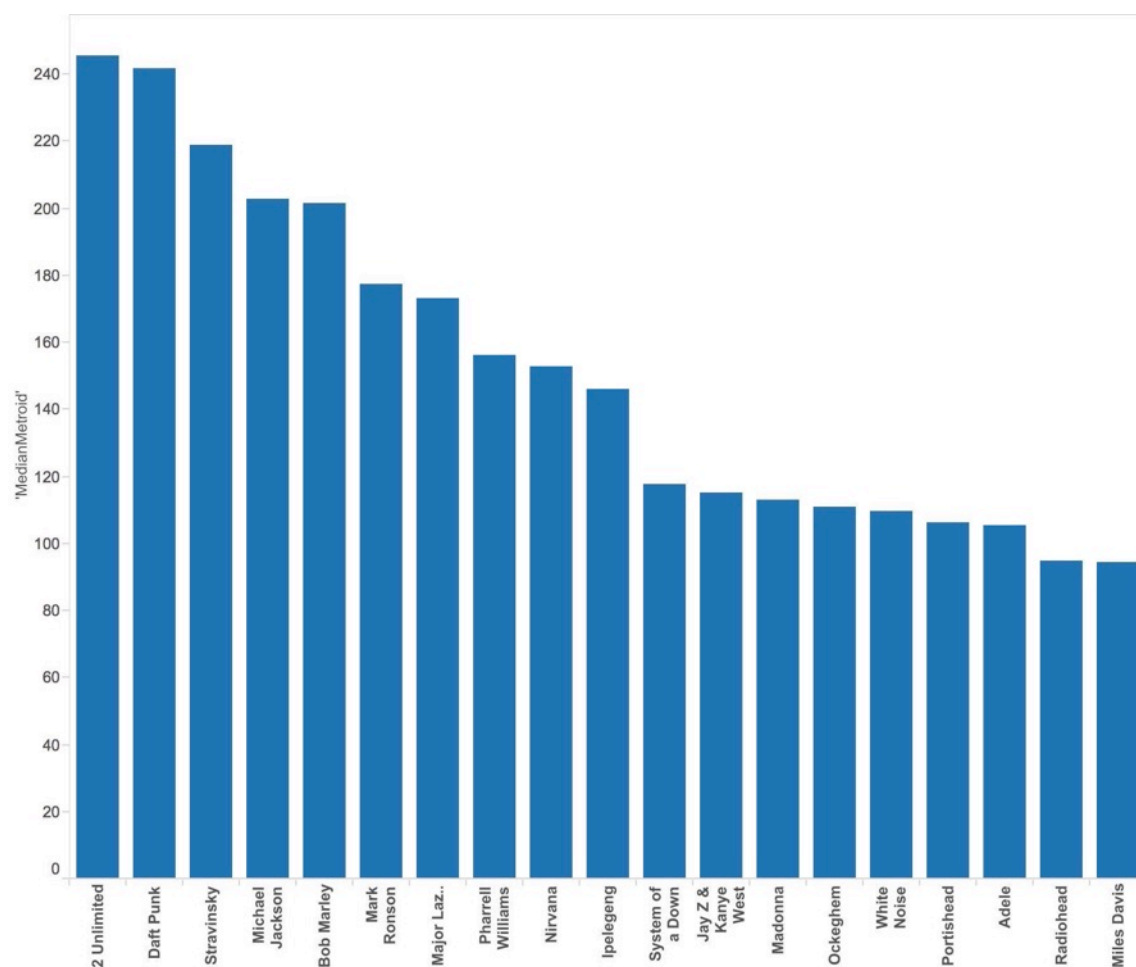


Figure 32 Median metroid values for the reference corpus.

A similar pattern appears when I investigate Metroid values for EDM music: Music with distinct subdivisions, mostly due to distinctive hi-hat or other markings of off-beats, score high Metroid values. For my dataset, I also created a measure for the *Metroid/Tempo*, which measures the amount of accentuation in between beats. And *MeanModeMetroid* which, like *MeanModeTempo*, represents the mean of each minute's mode of the Metroid.

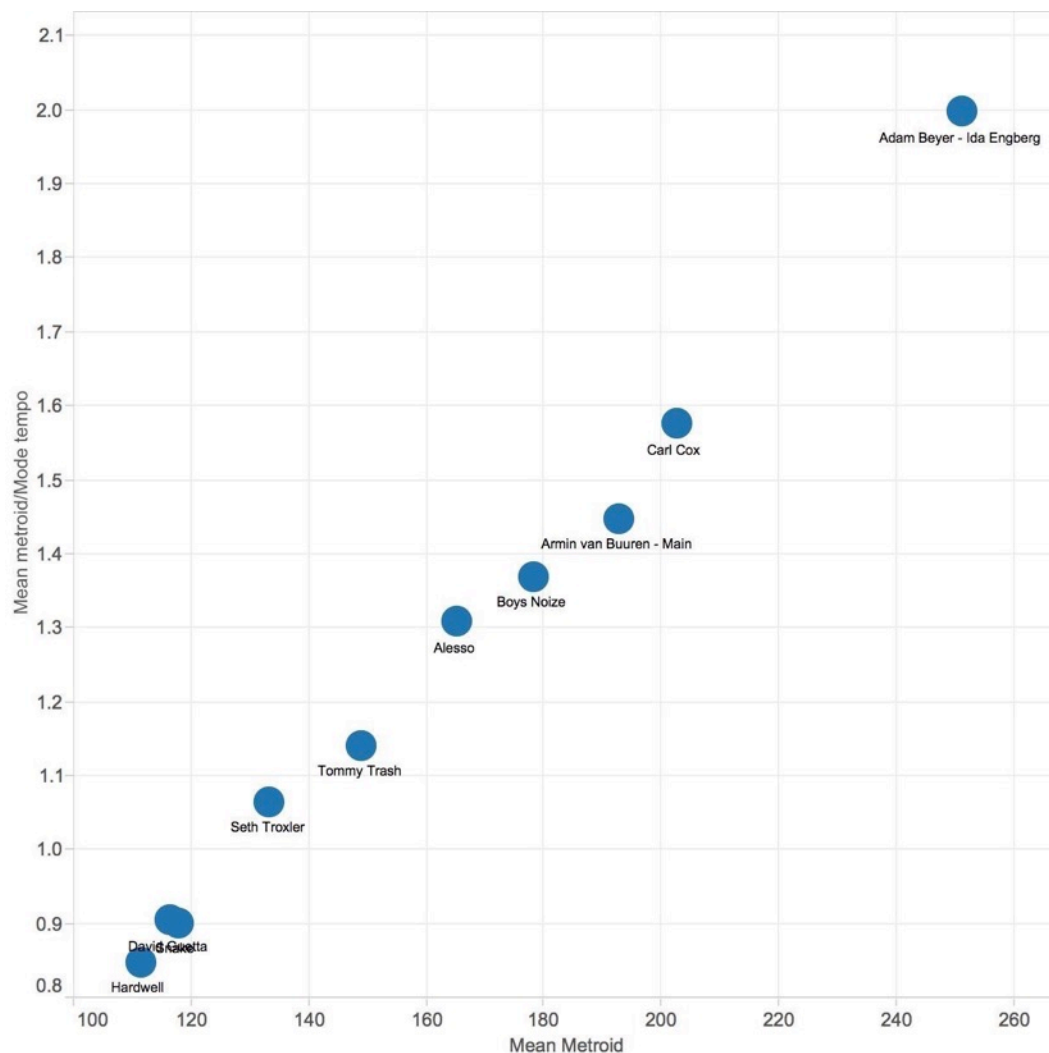


Figure 33 Mean metroid values plotted against and Mean metroid/tempo values for my test DJ subset. The measurements only account for minute 10-20 of these DJ sets.

Pulseclarity

Pulseclarity “[e]stimates the rhythmic clarity, indicating the strength of the beats estimated by the mirtempo function” (Lartillot 2014, 111; see also Lartillot et al. 2008b) Figure 34 displays a plot of the development of the reference corpus’ Pulseclarity values throughout the songs. Generally, the more prominent, distinctive and clear the percussive sounds are in the mix the higher the value. 2 Unlimited, Bob Marley, Daft Punk and Michael Jackson all have high Pulseclarity values throughout the songs; these songs have short and relatively high frequent bass drum hits, and other distinctive percussive sounds.

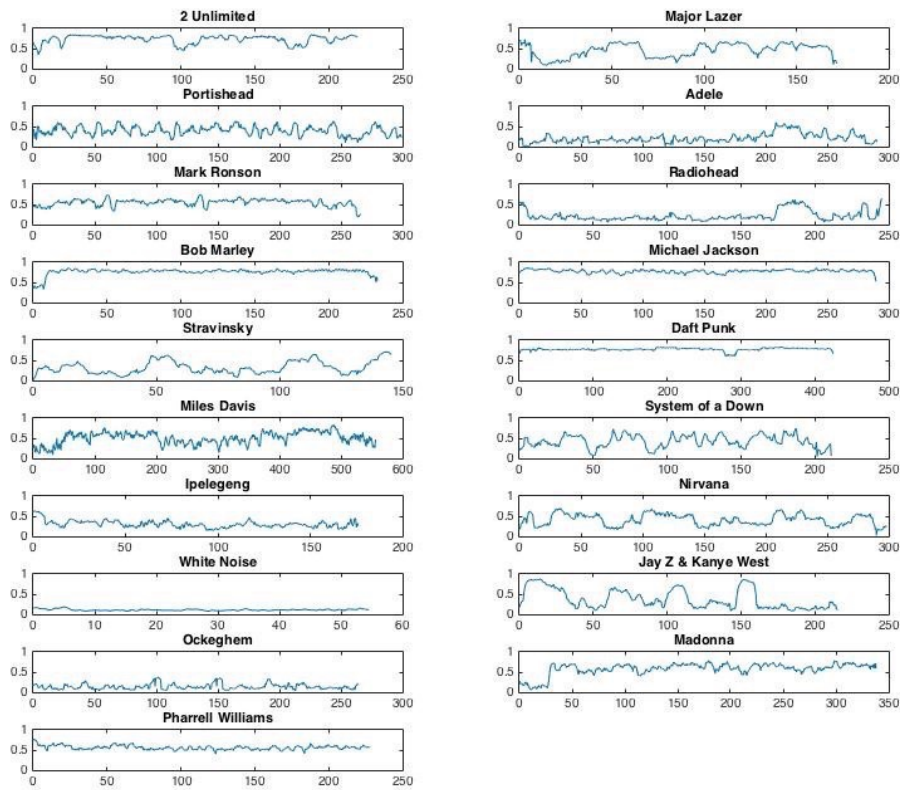


Figure 34 Pulseclarity values for the reference corpus as a function of time in seconds

It is very indicative, how and when Jay Z & Kanye West's *Ni**as In Paris* alternates between high and low Pulseclarity values: The short and high-pitched snare drum and synth from second 5-32 imply high Pulseclarity values. However, the very deep, long sounding and slightly distorted bass drum, which enters at 0.33, probably is the factor that causes the Pulseclarity value to drop to very low. This connection between sound and values recurs throughout the song. Especially from 2.47, Pulseclarity becomes very low. This is most likely a result of both the bass drum and a very distorted, long lasting sound that has the function of being a snare drum with a long release.

When plotting Pulseclarity for the DJ test corpus compositional traits of bass drum or no bass drum come into view. Seth Troxler's removing drums is very apparent and easily readable in low Pulseclarity values at 650 (10.50) and 780 (13.00). Around 960 (16.00) there is a lot of speech in the recording, due to an interview on top of the recording, and the Pulseclarity value drops to around 0.5. The 0.5 corresponds to the Pulseclarity values

of the sounds in the talking. In Hardwell, the passages without bass drums are also detectable from the plots. For example, around 610 (10.10-10.37) and 850-940 (14.10-15.40). However, in the more distorted passages, the music holds medium Pulseclarity values, which leaves it hard to read the presence of the beat directly from the graph.

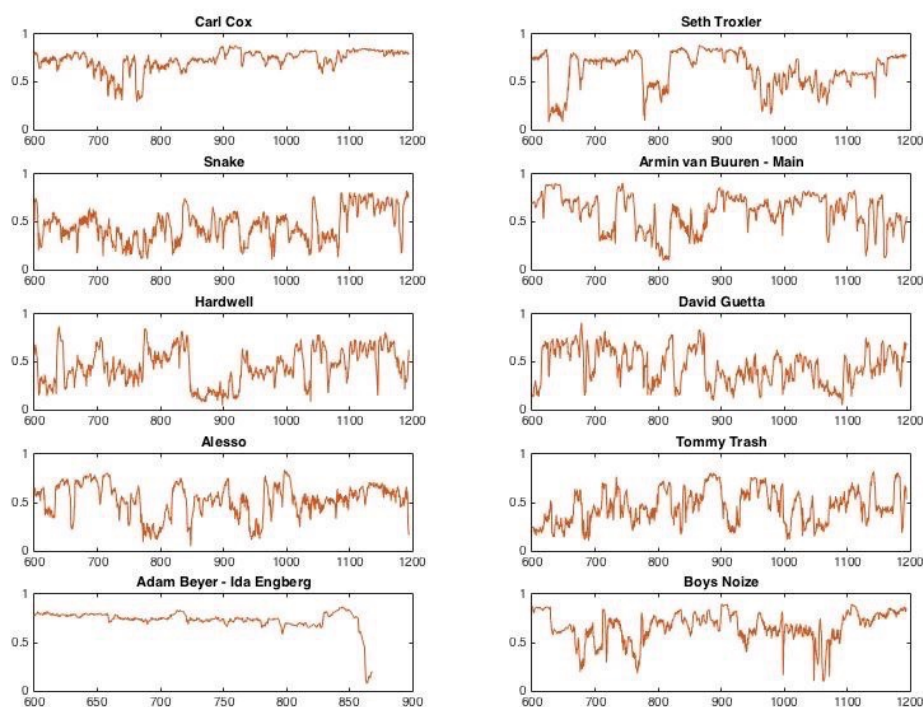


Figure 35 Pulseclarity values for 10 DJ subsets as a function of time in seconds.

Eventdensity

Eventdensity “[e]stimates the average frequency of events, i.e., the number of note onsets per second” (Lartillot 2014, 98)

In the reference corpus, not surprisingly the long drawn out sound in Ockeghem result in low Eventdensity level. Just as the development of Adele’s Hello becomes apparent here; in the intensive passages Adele sings more notes leading to higher Eventdensity value. System of a Down’s alternation between fast paced, distorted sections and quiet ones also is reflected in these values. Interestingly the many fluctuations in White Noise are considered events by this feature, White Noise has by far the highest Eventdensity value.

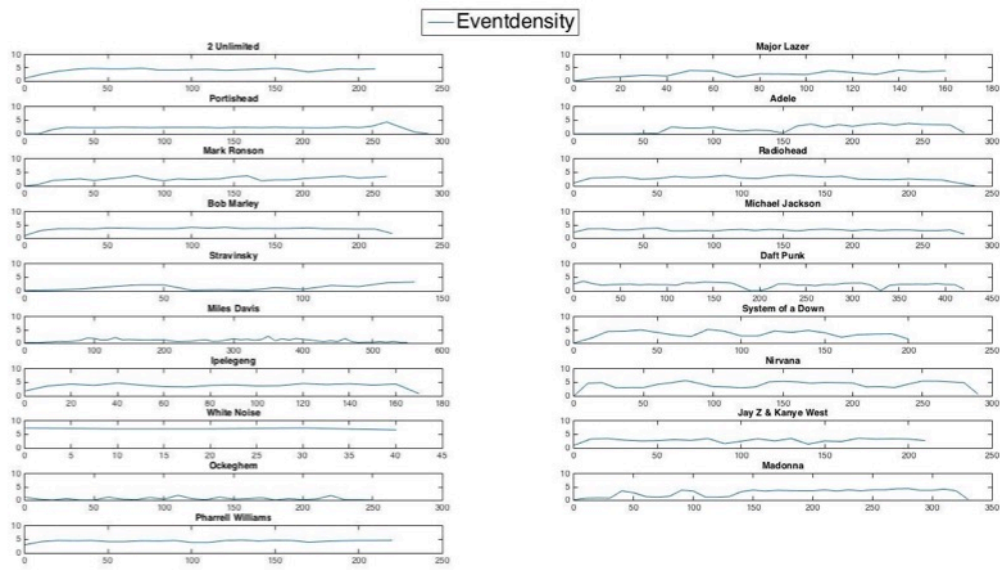


Figure 36 Eventdensity values for the reference corpus as a function of time in seconds.

For the DJ test corpus, the two breaks identified in the pulseclarity values of Seth Troxler measure have different sonorous characteristics. These different characteristics are identifiable in Eventdensity. The first at 620 (10.20) contains drawn out sounds resulting in low Eventdensity values. While the other break, though it lacks bass drum, contains more events, mainly more staccato sounds.

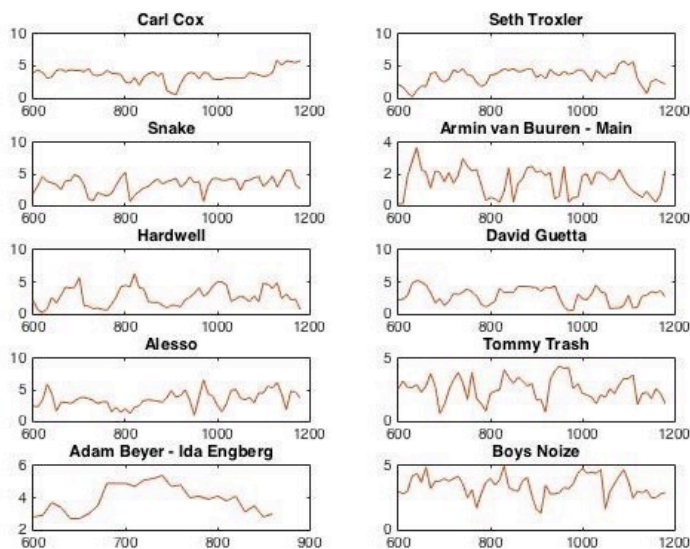


Figure 37 Eventdensity values for the DJ test corpus as a function of time in seconds.

7.2.3 Timbral Features

The timbral features are all measures that measure aspects of the sound signal rather directly:

Zerocross and Roughness estimate aspects of discordance; noise and dissonance: More precisely, Zerocross is “[a] simple indicator of noisiness consists in counting the number of times the signal crosses the X-axis (or, in other words, changes sign)” (Lartillot 2014, 123). While Roughness is “[a]n estimation of the total roughness [...] by computing the peaks of the spectrum, and taking the average of all the dissonance between all possible pairs of peaks” (133).¹⁴⁴

Brightness and Rolloff measure aspects of the audio files’ frequency balance: Brightness measures the amount of energy above 1500 Hz (127). Rolloff estimates the frequency in which 85% of energy is below and 15% above (125).

Irregularity and Flux measure aspects of the variations in timbre: Flux describes the “distance between successive frames” in the spectrum (60). While “Irregularity of a spectrum is the degree of variation of the successive peaks of the spectrum” (135). Both are calculated for frames of 50 milliseconds.

The timbral features will be examined more closely in the next Section 7.3, in which I have access to data on how much they mutually correlate.

7.2.4 Tonal features

In MIRtoolbox, as with ACA methods in general, tonal features are computed from the chromagram, which is computed from the spectrum. The Chromagram “shows the distribution of energy along the pitches or pitch” (Lartillot 2014, 145)¹⁴⁵, and is consequently an indicator of which pitches that sound at given time. The pitches can be constricted and summarized into chroma classes, by calculating the energy of each tone in the chromatic scale, as depicted in Figure 38.

¹⁴⁴ Referencing (Sethares 1998)

¹⁴⁵ (Müller et al. 2011; Müller 2015) explain well how pitch values can be calculated from the spectrum.

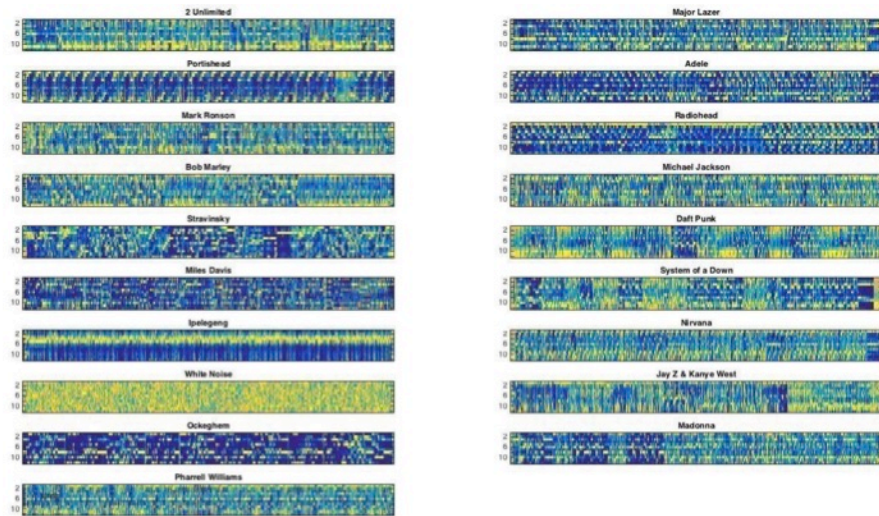


Figure 38 Chromagrams for the reference corpus.

One of the major differences, compared to score analysis, is that musical content that previously was not considered content to be analyzed as tonal content now becomes included because it sounds with a pitch. Percussive elements or overtones bring for example tonal content along which becomes included in the measurements.

For my analysis, I apply Mode, the Harmonic Change Detection Function (HCDF), and Inharmonicity.

Mode

Mode “[e]stimates the modality, i.e. major vs. minor, returned as a numerical value between -1 and +1: the closer it is to +1, the more major the given excerpt is predicted to be, the closer the value is to -1, the more minor the excerpt might be” (Lartillot 2014, 155). As with Echo Nest, these values indicate probabilities of either one of the two modes; major or minor mode (see Lartillot 2014, 149 for explanation). The measurements are very local; a song overall written in minor mode, but containing more major chords is likely to

attain more positive than negative Mode values.

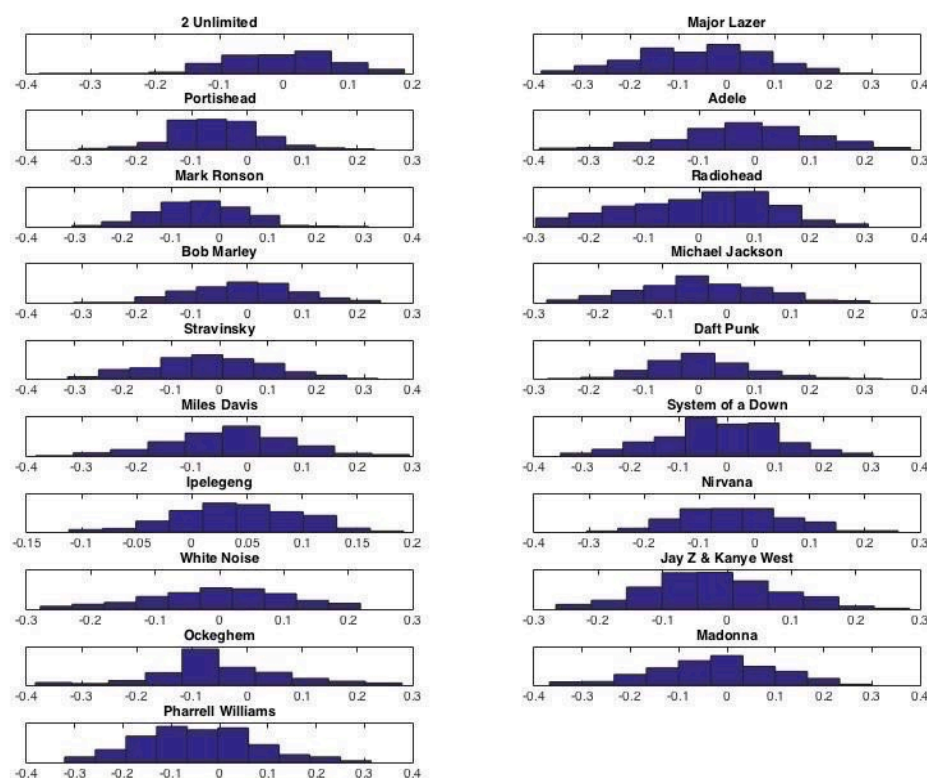


Figure 39 Histogram of the Reference corpus' Mode values

HCDF and Inharmonicity

“The Harmonic Change Detection Function (HCDF) is the flux of the tonal centroid” (160) which “corresponds to a projection of the chords along circles of fifths, of minor thirds, and of major thirds” (159; Harte et al. 2006).

HCDFoverI

HCDF measurements are calculated for intervals of 743 ms, but I found it complicated to interpret the relationship between the measurements and tonal change. Precisely which aspects of tonal change are measured? I therefore decided to compute a new measure, *HCDFoverI*.¹⁴⁶ See definition below.

¹⁴⁶ `getHCDFdata.m`

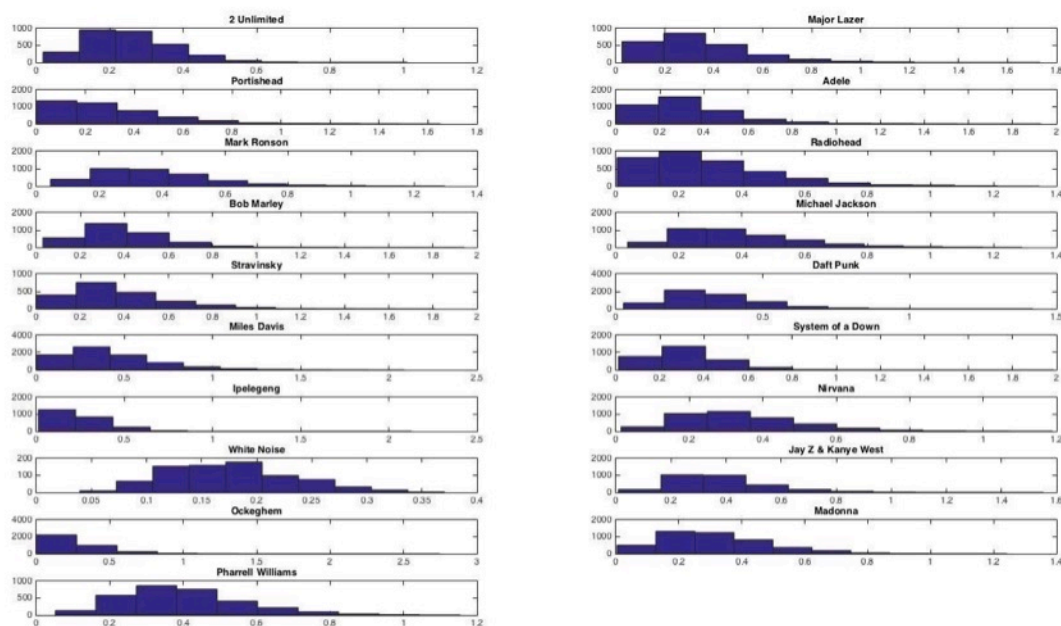


Figure 40 Histogram of HCDF-values for the reference corpus. Pay attention to the X axes as they indicate the highest HCDF-value for each song.

HCDFover1 is a measure of the number of frames per 10 seconds that have a HCDF value over 1. The number indicates the amount of more profound harmonic changes relative to the circle of fifths.

Inharmonicity

Inharmonicity estimates “the amount of energy outside the ideal harmonic series.” The feature is based on “a simple function estimating the inharmonicity of each frequency given the fundamental frequency f_0 ”. This “simple model presupposes that there is only one fundamental frequency” (Lartillot 2014, 143). In other words, Inharmonicity is a simple indicator of dissonance, but again the relation between measurement and music is opaque. The reference corpus provides some indicators: Songs with harsh, distorted, electrified, percussive sonorities score high, while songs with more mellow sounds with fewer overtones score low.

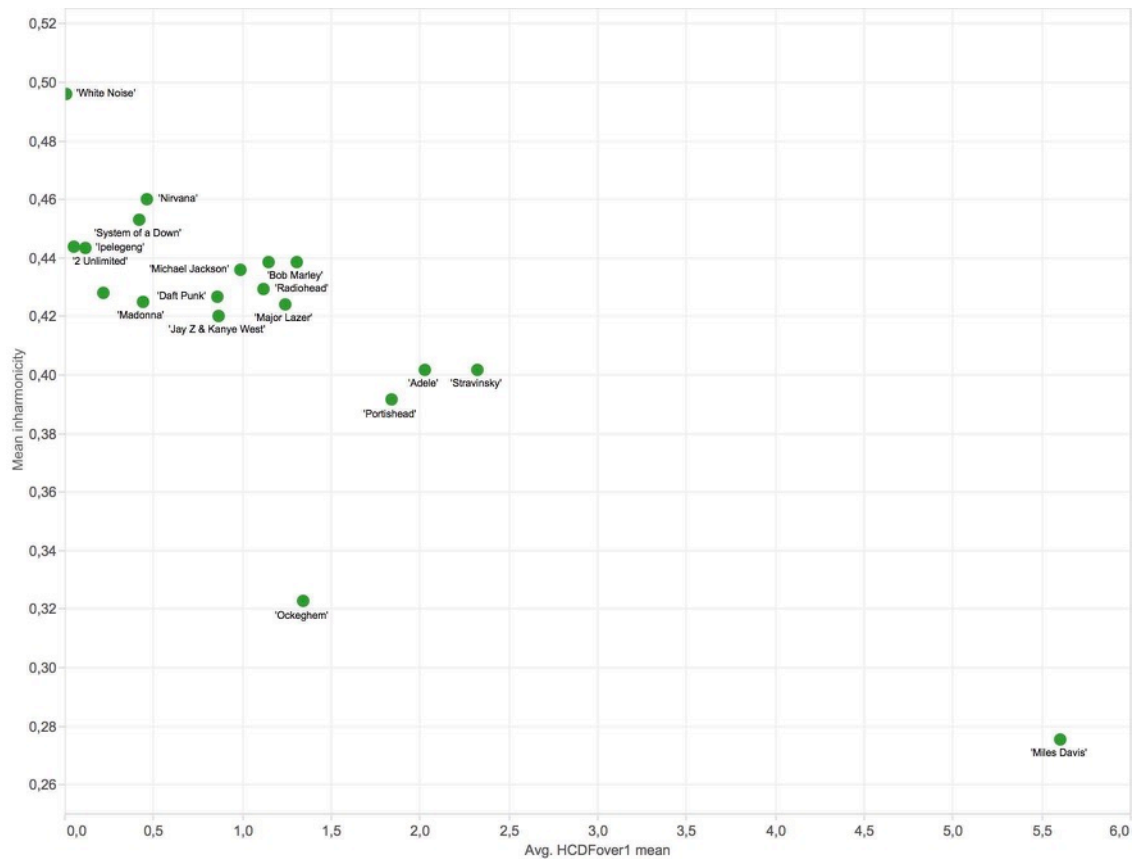


Figure 41 HCDFover1 and Inharmonicity for the reference corpus.

In Figure 41 above, I plotted Inharmonicity against HCDFover1 values. The songs that change harmonically more profound hold high HCDFover1 values. However, again these changes only concern the level of seconds. For example, Portishead holds a high HCDFover1 value despite that the song overall is fairly tonally settled as it is based on a repeated 2 bar loop. Nevertheless, one or two of the chord changes within the loop are assigned an HCDF value over 1, and therefore the HCDFover1 value becomes almost 2, implying almost two profound harmonic changes per 10 seconds.

7.2.5 A general remark on post processing

It applies to all features that they are dependent on aspects of the sound that they are affected by the processing from the sounding music to the audio file. Other agents, such as sound engineers, recording techniques, the audio system, audio compression formats, etc. can affect the sound of the final audio file. On the one hand, these aspects are largely irrelevant in relation to investigating DJs' music choices: They blur the relationship

between the features and the questions I pose. If a sound engineer chooses to amplify the 4 kHz band, it will affect the features, especially the timbral ones, compared to if the band was not amplified. On the other hand, a DJ's preferences' for music and certain types of timbre might be detectable in the timbral features.

7.3 Step 2 - Surface Views: Exploring the Datasets by Mapping Them

After having investigated the features' relations to the music, the next step was to explore my corpus with them. For that purpose, I calculated the mean values for each feature for each audio file.¹⁴⁷ I discarded audio files shorter than 30 seconds. My first step was to deploy a principal component analysis (PCA) for the initial structuring of the corpus, and for automatically searching for patterns in the quantitative metadata; I wanted to create a visual map of the audio files and to investigate the features' mutual connection.¹⁴⁸

Though I collected other statistics about each audio file, I chose to map only mean values, to reduce the complexity of the findings. One problem with statistical values such as standard deviation that describe variations in the dataset is that they concern both deviations from the musical level of milli-seconds to the length of the full audio file. Therefore their connection to the music is unclear. Mean values, coarsely described, represent somewhat average textures¹⁴⁹ of the music. However, one problem with mean values is that they do not take musical development such as contrasting passages into account. Instead, they average contrasts out.

7.3.1 The features

I created a principal component analysis from features that are normalized due to standard deviation and averages.¹⁵⁰ Principal component 1 and 2 can describe up to 26 percent¹⁵¹ of the overall variance in the dataset. Hence, a 2D plot represents about a fourth of the

¹⁴⁷ I applied the script, `getIntervalData.m`, which was set up also to calculate mean, median, standard deviation and inter quartile range. See '480DJs subsets' for a full dataset containing all statistics of each audio file.

¹⁴⁸ The script `getAllStats.m` accounts for the process of collecting the stats and creating PCA view of them.

¹⁴⁹ Oxford Music Online (Newbould 2017) writes about *texture*: "If a series of snapshots could be taken, in fairly quick succession, of the vertical cross-section of a musical passage, these might provide a basis for determining the texture of the music. [...] Texture thus describes the vertical build of the music—the relationship between its simultaneously sounding parts—over a short period of time." This corresponds very well to how ACA methods measure on music.

¹⁵⁰ `PCAjesper.m`

¹⁵¹ PCA 1-4 describes up to 46 percent.

complexity of the full dataset. Nevertheless, it provides a lot of indications of overall tendencies in the dataset. In this first part, I will discuss what we can learn about the features from investigating the loadings (Figure 42 and Figure 43).

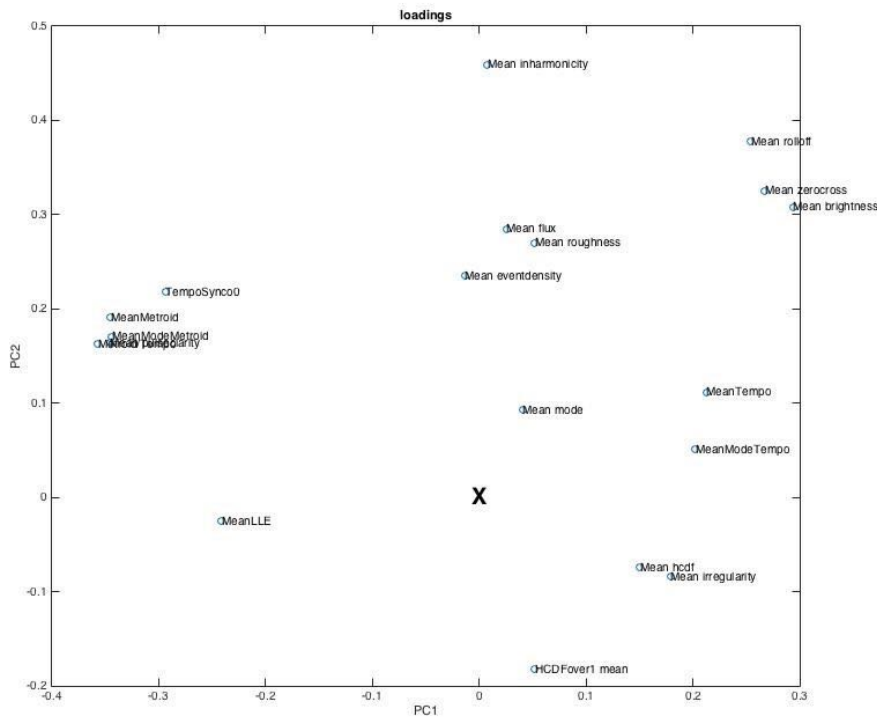


Figure 42 Loadings plot of PC1 and PC2 for 480 audio files containing more than 30 seconds of sound.

For 480 DJ excerpts	Mean																		
	Mean zero-cross	Mean rolloff	Mean brightness	Mean roughness	Mean irregularity	Mean flux	Mean event density	Mean LLE	Mean Metroid	Mean Mode Metroid	Mean pulseclarity	Mean tempo	Mean ModeT	Mean mode	Mean hcdf	Mean inharmonicity	TempoSynco	HCDFOver1 mean	Metroid/Tempo
Mean zero-cross	1,00	0,80	0,93	0,05	0,03	-0,03	0,07	-0,31	-0,43	-0,44	-0,46	0,28	0,25	0,12	0,18	0,33	-0,29	0,04	-0,45
Mean rolloff	0,80	1,00	0,87	0,08	0,21	0,12	0,07	-0,31	-0,36	-0,37	-0,39	0,30	0,25	0,13	0,25	0,39	-0,21	0,00	-0,39
Mean brightness	0,93	0,87	1,00	-0,01	0,16	-0,03	0,04	-0,37	-0,48	-0,49	-0,52	0,32	0,28	0,15	0,24	0,34	-0,33	0,03	-0,50
Mean roughness	0,05	0,08	-0,01	1,00	0,21	0,91	0,28	-0,18	0,00	-0,03	-0,07	0,15	0,08	-0,04	0,10	0,08	-0,12	-0,02	-0,02
Mean irregularity	0,03	0,21	0,16	0,21	1,00	0,25	0,21	-0,45	-0,29	-0,30	-0,42	0,20	0,20	0,03	0,32	-0,25	-0,38	0,15	-0,31
Mean flux	-0,03	0,12	-0,03	0,91	0,25	1,00	0,22	-0,13	0,07	0,05	-0,01	0,12	0,04	-0,02	0,09	0,08	-0,03	-0,06	0,04
Mean event density	0,07	0,07	0,04	0,28	0,21	0,22	1,00	-0,23	0,15	0,13	0,11	-0,08	-0,07	-0,02	0,13	0,07	0,11	0,02	0,16
MeanLLE	-0,31	-0,31	-0,37	-0,18	-0,45	-0,13	-0,23	1,00	0,42	0,44	0,49	-0,31	-0,26	-0,04	-0,22	0,06	0,44	-0,04	0,44
MeanMetroid	-0,43	-0,36	-0,48	0,00	-0,29	0,07	0,15	0,42	1,00	0,96	0,78	-0,32	-0,31	-0,05	-0,28	0,08	0,63	-0,10	0,99
MeanModeMetroid	-0,44	-0,37	-0,49	-0,03	-0,30	0,05	0,13	0,44	0,96	1,00	0,77	-0,33	-0,30	-0,03	-0,26	0,06	0,63	-0,08	0,95
Mean pulseclarity	-0,46	-0,39	-0,52	-0,07	-0,42	-0,01	0,11	0,49	0,78	0,77	1,00	-0,32	-0,33	-0,08	-0,33	0,16	0,80	-0,14	0,79
MeanTempo	0,28	0,30	0,32	0,15	0,20	0,12	-0,08	-0,31	-0,32	-0,33	-0,32	1,00	0,80	0,04	-0,02	0,10	-0,36	-0,11	-0,43
MeanModeTempo	0,25	0,25	0,28	0,08	0,20	0,04	-0,07	-0,26	-0,31	-0,30	-0,33	0,80	1,00	0,05	0,03	0,00	-0,37	-0,07	-0,40
Mean mode	0,12	0,13	0,15	-0,04	0,03	-0,02	-0,02	-0,04	-0,05	-0,03	-0,08	0,04	0,05	1,00	0,01	0,07	0,00	-0,05	-0,06
Mean hcdf	0,18	0,25	0,24	0,10	0,32	0,09	0,13	-0,22	-0,28	-0,26	-0,33	-0,02	0,03	0,01	1,00	-0,10	-0,24	0,55	-0,27
Mean inharmonicity	0,33	0,39	0,34	0,08	-0,25	0,08	0,07	0,06	0,08	0,06	0,16	0,10	0,00	0,07	-0,10	1,00	0,25	-0,17	0,06
TempoSynco	-0,29	-0,21	-0,33	-0,12	-0,38	-0,03	0,11	0,44	0,63	0,63	0,80	-0,36	-0,37	0,00	-0,24	0,25	1,00	-0,14	0,64
HCDFOver1 mean	0,04	0,00	0,03	-0,02	0,15	-0,06	0,02	-0,04	-0,10	-0,08	-0,14	-0,11	-0,07	-0,05	0,55	-0,17	-0,14	1,00	-0,08
Metroid/Tempo	-0,45	-0,39	-0,50	-0,02	-0,31	0,04	0,16	0,44	0,99	0,95	0,79	-0,43	-0,40	-0,06	-0,27	0,06	0,64	-0,08	1,00

Figure 43 Correlation values between features¹⁵². Values higher than 0.75 marked with red.

¹⁵² Calculated with MATLAB's corr-function

These figures above are able to demonstrate general stylistic trends found in the dataset. In Figure 42 alongside the x-axis, which represents the first principal component¹⁵³ are the rhythmical features: Metroid is placed far from the X¹⁵⁴ to the left, while Tempo is positioned in the opposite direction. Musically, this indicates that within the dataset there is a general tendency that slower tempo implies stronger accentuations in between beats.¹⁵⁵ There is also a tendency that lower tempo implies averagely more “space” between beats, because of high correlation between Metroid and both LLE values and Pulseclarity values. On the right side of the plot, high tempo correlates to some extent with Brightness, Zerocross, and Rolloff, but also with Roughness and Irregularity. This indicates a general tendency that audio files in the corpus with higher Tempo also are more noisy and dissonant.

The data also shows a tendency that low tempo implies higher tempo confidence, measured in TempoSynco. This opposes a general trend in popular music that slower, rhythmically focused music tend to involve more syncopation. But within this dataset, higher tempo implies lower TempoSynco. This may be due to the fact that the vast majority of the music has four-on-the-floor on the bass drum. And the few DJs that do not tend to prefer this type of rhythm to some extent fall into the double tempo problem. The implication is that music a human would have categorized as low tempo-high syncopation is estimated by MIRtoolbox as high tempo-high syncopation. DJ Snake and Skrillex are examples of this problem.

Tonally, the usage of either major or minor mode does not correlate very well with any of the other features. Graphically, this is shown by mode being placed closed to the X, which represents 0.0. Hence, there are apparently no clear stylistic trends in relation to the usage of either major or minor mode that can be identified through the dataset. Correspondingly with the new measure of harmonic change, HCDFover1, which does not correlate significantly with any other features. However, mean HCDF correlates somewhat with bright sounds and features that indicate discordance. A possible explanation could be that HCDF becomes affected by noise, which in combination with more high pitched sounds send energy to more notes in the chromagram. This dispersion of tonal energy into more pitch classes could imply insecurity on tonal issues.

It is somewhat surprising that Roughness and Inharmonicity do not correlate very well (0.08), despite that both features indicate aspects of dissonance. However, if applied only

¹⁵³ Accounting for 16% of the variance in the dataset.

¹⁵⁴ Which displays 0,0

¹⁵⁵ The correlation matrix confirms this, because tempo and metroid correlates with -0.32.

to the reference corpus¹⁵⁶ the correlation value between these features is 0.88.¹⁵⁷ (See Figure 44). This disparity occurs relevant and can potentially illuminate both properties of the music and the features. However, I found both features too complicated to account for and therefore I am not able to provide a music analytic reason for why this discrepancy occurs.

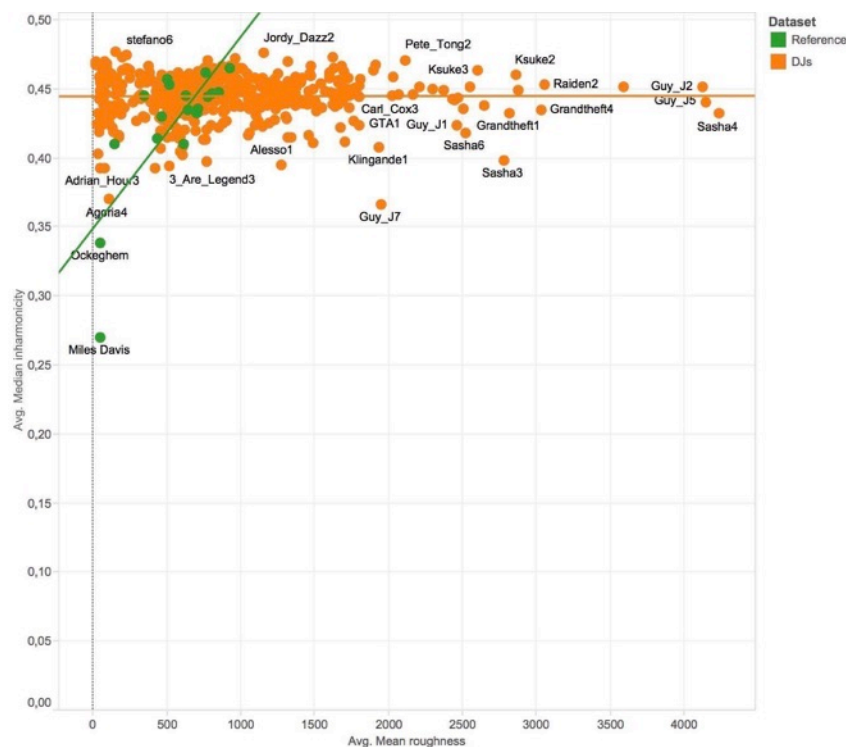


Figure 44 Inharmonicity and Roughness values for the DJs and reference corpus, including trend lines for both corpora.

The correlation matrix shows two groups of correlating timbre features: Brightness correlate with Rolloff and Zerocross, and Roughness and Flux correlate strongly. Irregularity is somewhere in between these groups, connecting a little to both groups.

Timbrally, it is no big surprise that Brightness and Rolloff correlate a lot because they both concern the distribution of spectral information; the degree of bright sounds predominate. However, that Zerocross correlate strongly with these two features is far from self-evident. One reason could be that high frequencies oscillate more, causing higher Zerocross values. Notwithstanding, the music analytical implication is that confusion and

¹⁵⁶ White Noise excluded

¹⁵⁷ getAllStats.m and 480DJs subsets > Correlation Matrixes > #2

doubt arise: Do these features measure mostly noise and distortion *or* the distribution of frequencies?

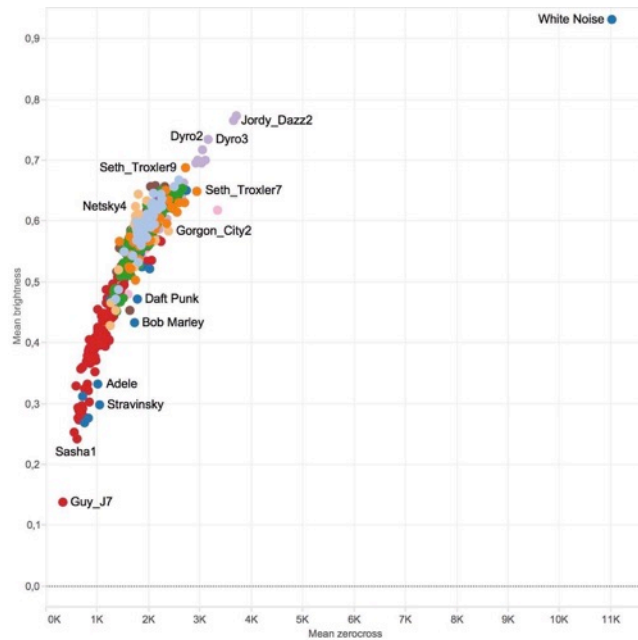


Figure 45 Zerocross and Brightness values for both the DJs and reference corpus.

It is neither self-evident that Roughness and Flux correlate as much as 0.91. As also indicated in Figure 46.

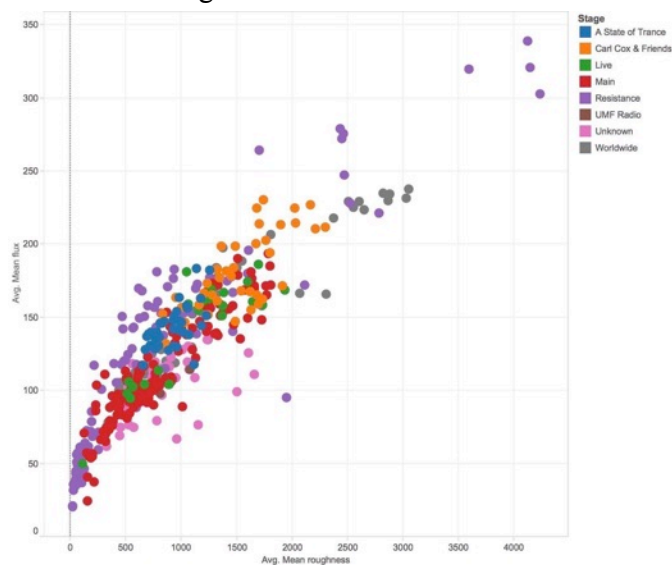


Figure 46 Mean Roughness and mean Flux values for both the DJs and reference corpus.

A plot of these two features over the course of the whole songs in the reference corpus provide insight in how they relate, and what they measure:

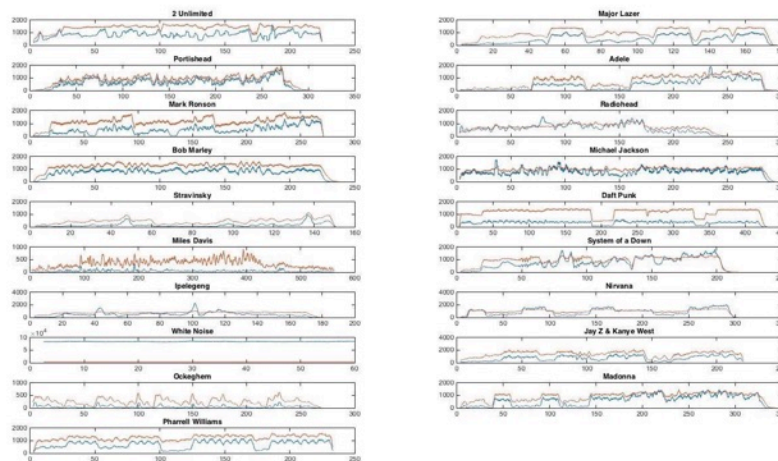


Figure 47 Flux x 10 (red) and Roughness (blue) of the songs in the reference corpus. X-axis represents time in seconds. The lines display moving averages for intervals of 2.5 seconds.

The musical development is especially apparent in the plot of Pharrell Williams, where the sections with few instruments, claps and vocal harmony imply low Roughness and Flux values. The same pattern occurs in Madonna: The quiet passages in the verse hold low Roughness and Flux, while the chorus is louder and much more energetic by including more instruments amongst other drums, bass, and guitar. Consequently, the chorus also has higher Roughness and Flux values than the verse. Interestingly, White Noise has low Flux but very high Roughness. This demonstrates a general tendency that distorted music has low Flux/Roughness ratio. This is further exemplified especially in the chorus of Nirvana, where the blue line exceeds the red.

7.3.2 The music

The other part of the PCA is the plot of the music, relating to the loadings' plot above. This a good place to start the investigation of the corpus.

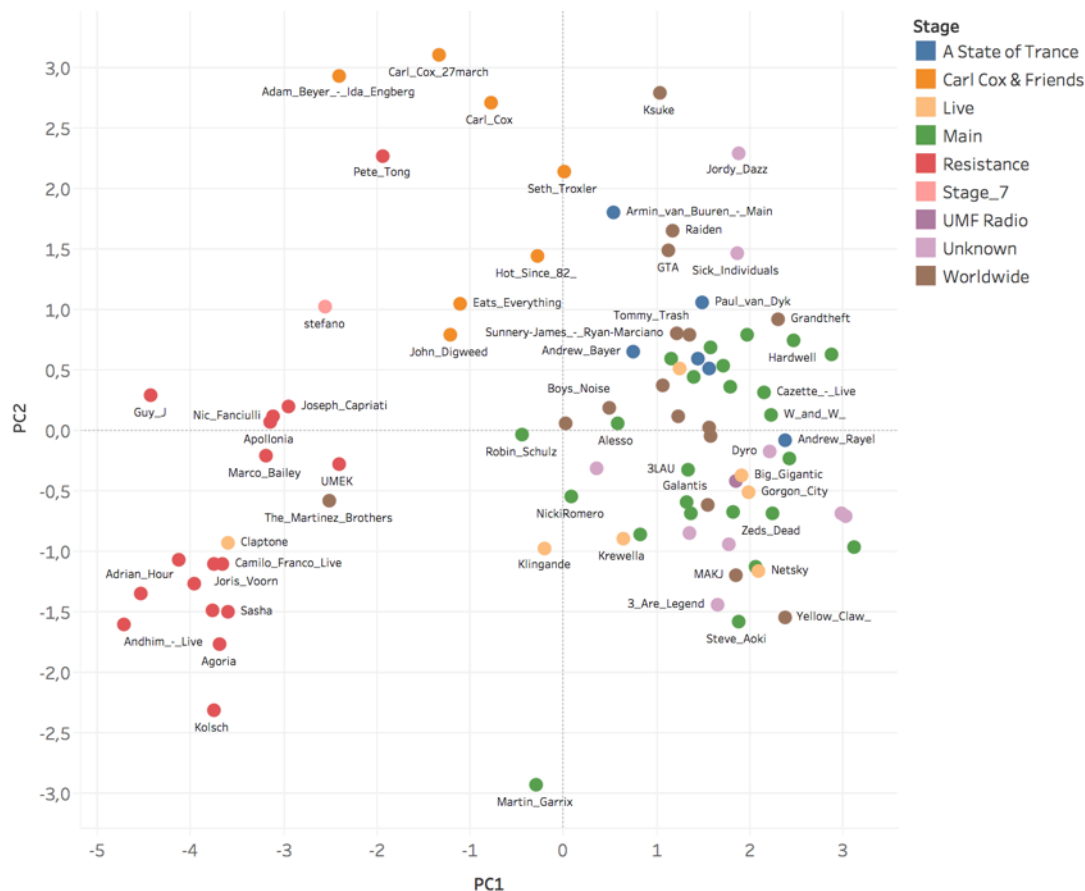


Figure 48 Average PC1 and PC2 of the audio files from each act in the dataset. The colors represent the stage they played at. See Figure 42 above for the loadings' plot.

The two axes represent a complex combination of features. However, in the main, the plot primarily illustrates which acts hold similar features. It becomes clear that the music played at different stages is plotted nearby each other more or less in clusters. This demonstrates that this method to a large extent can distinguish between the music played at the various stages. And it is a strong indicator that the music played at the different stages represents different subgenres within EDM. This is confirmed through listening. When compared to the loadings' plot, it becomes apparent that the music played there are three main groups: Resistance stage's music is generally slower than the other stages', and have higher Pulseclarity, LLE, and Metroid. The music at Carl Cox & Friends Stage has higher Pulseclarity and Inharmonic. While Main Stage and State of Trance Stage are closely related in the plot, probably due to the usage of similar sonorities, amongst other a lot of sawtooth wave based synth sounds.

But the PCA plot displays a compressed view, representing only 26% of the variance in the dataset, and it is necessary to consult the statistics to verify the tendencies that are suggested:

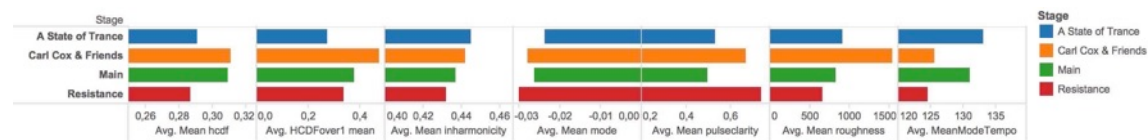


Figure 49 Average statistics for four stages.

The statistics indicate that the music at Carl Cox & Friends tend to change chords more often, despite being located at the top, while the music is more dissonant than the other stages', indicated by higher Roughness and Inharmonicity values. Resistance stage is most minor-ish, has the lowest tempo, and the least dissonance. State of Trance and Main Stage play the fastest music, approximately 5-10 beats per minute faster than the other stages in average.

The properties of EDM - zooming further in and out

Despite the diversity displayed in the plot, the DJsets comprise a rather homogeneous corpus, when compared to other genres. Figure 50 shows the degree of differences between the acts playing at the four stages, compared to the reference corpus in the bottom:

The lack of stylistic diversity compared to the reference corpus' is reflected in more feature variation amongst the reference corpus. The PCA plot of all DJ subsets and the reference corpus in Figure 51 further confirms this relation. The statistics show what is rather obvious from listening to the music. The music played at the festival generally have lower HCDF values, which indicate more tonal modality, due to repeating loops staying within one tonality within EDM. The tempo of the festival is very uniform, for the major part between 125-135 BPM. And high Pulseclarity values indicate an emphasis on the beats comparable to the rhythmically focussed songs in the reference corpus. The tendency, also indicated in the PCA plot, manifest itself further in Figure 50's more detailed view of the features; audio files from the same stage resemble each other feature wise.

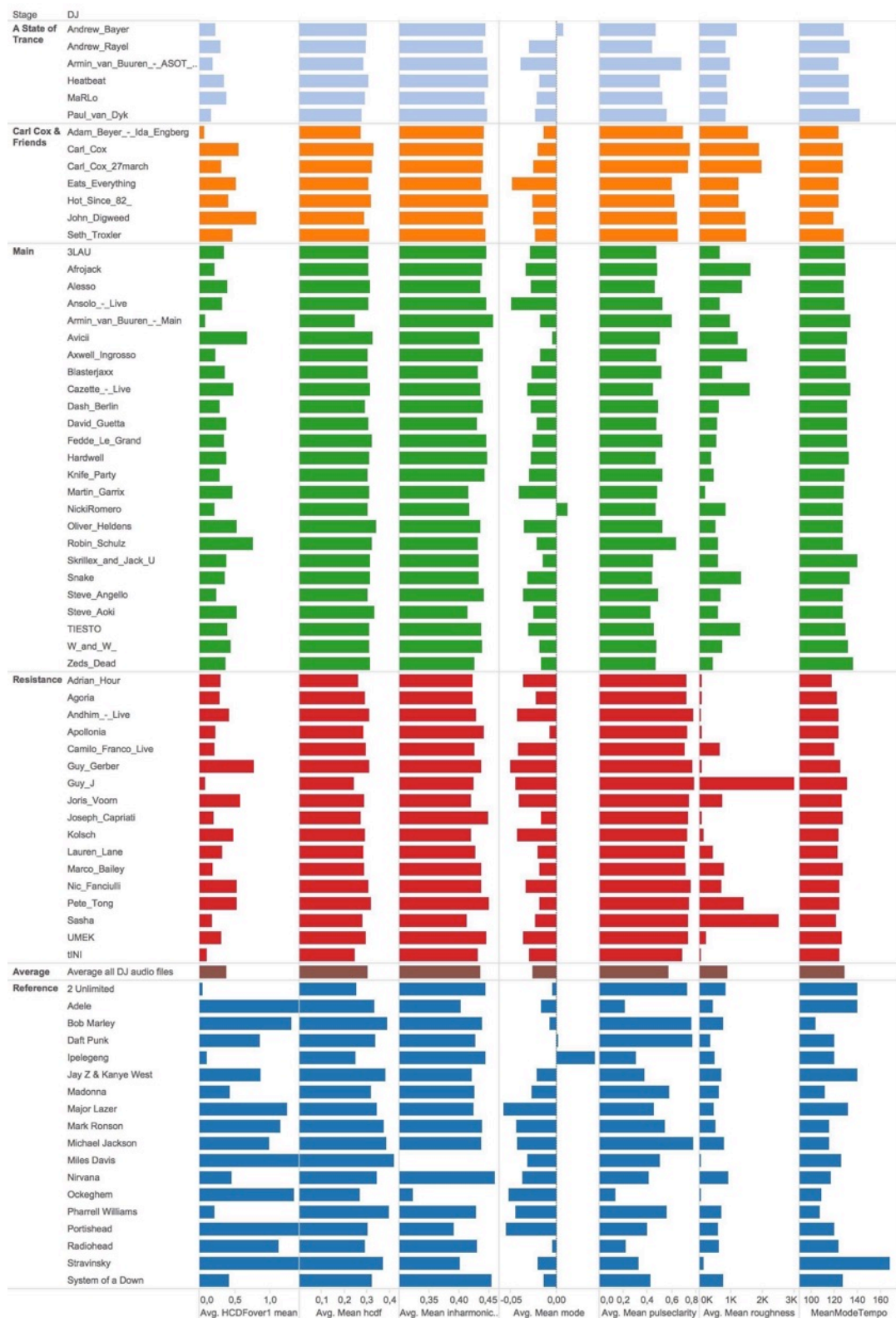
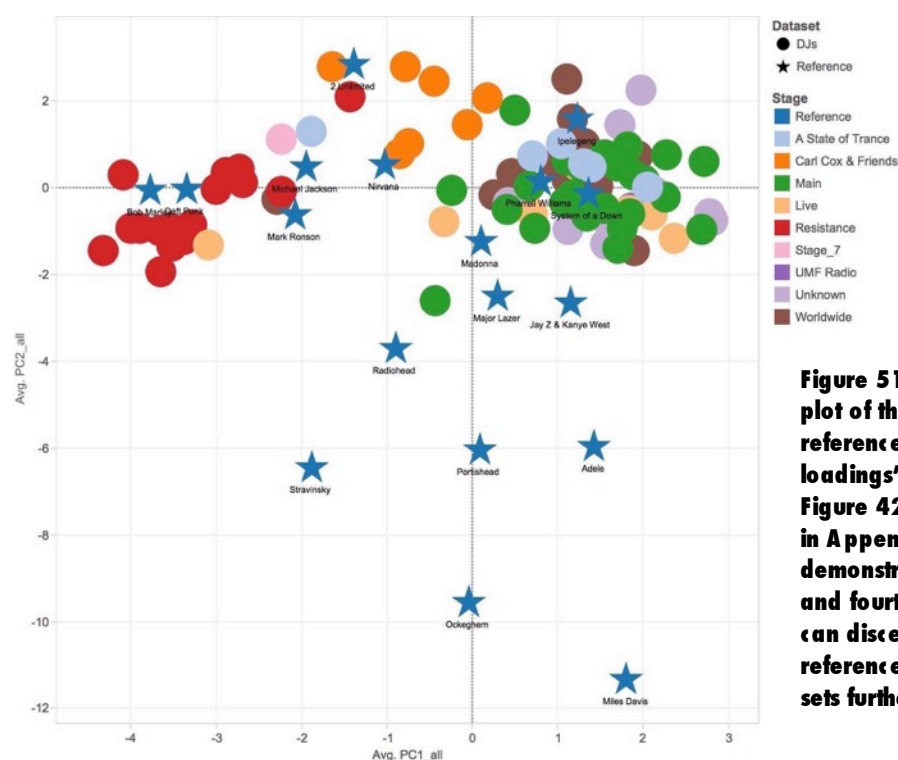


Figure 50 Average features for the music played at four stages compared to the reference corpus.

The datafication of the audio files thus enables me to go from the larger overview down to nuancing variations amongst singular instances. I can generalize about genres and subgenres with knowledge about local variations within these. And I can also find and choose representative examples with knowledge on their stylistical position within these: If I choose Avicii as a representative of EDM, he is not indicative of a general approach to tonal issues. Rather Avicii represents a link between EDM and mainstream, commercial music, and if I had collected data from chart music, this data could indicate, which musical aspects Avicii have in common with mainstream music, and which aspects he has most in common with EDM. The reference corpus already hints in a plausible direction: The sonorities and the sound idiom is closely related to much of EDM, while the harmonic progressions tend to involve more function harmonic chord progressions. Both these aspects are expressed in the statistics.

But there are also pitfalls in this way of viewing the music. For example, Guy J and Sasha stand out on Roughness. This may be due to bad sound quality. In both cases, the low frequencies are distorted, which may cause their high Roughness values, and consequently also their position on the PCA plot. Another issue that becomes apparent is content. In John Digweed's high HCDFover1 value is mostly due to an interview that starts at the beginning of 30.00 and influences the fourth and final part of the set profoundly. Without this fourth part, the HCDFover1 value would have been 0.4, equivalently to the other sets



at the stage. The data view can help identify these irregularities.

Figure 51 The PC 1 and PC 2 plot of the both DJ's and reference corpus. The loadings' plot resembles Figure 42, and can be found in Appendix 3. PC 3 and 4 demonstrate that the third and fourth PC component can discern the songs in the reference corpus from the DJ sets further, see Appendix 4.

7.4 Step 3 - The Shape of the Set: Analysis on the Macro level

7.4.1 The journey

It is often proposed that the DJ takes the audience on a journey and that DJs shape their sets accordingly. The notion of the journey thus concerns long term planning of the whole course of the set. Broughton & Brewster (2003, 135) have plotted archetypical strategies of shaping the set, regarding energy as a function of time.

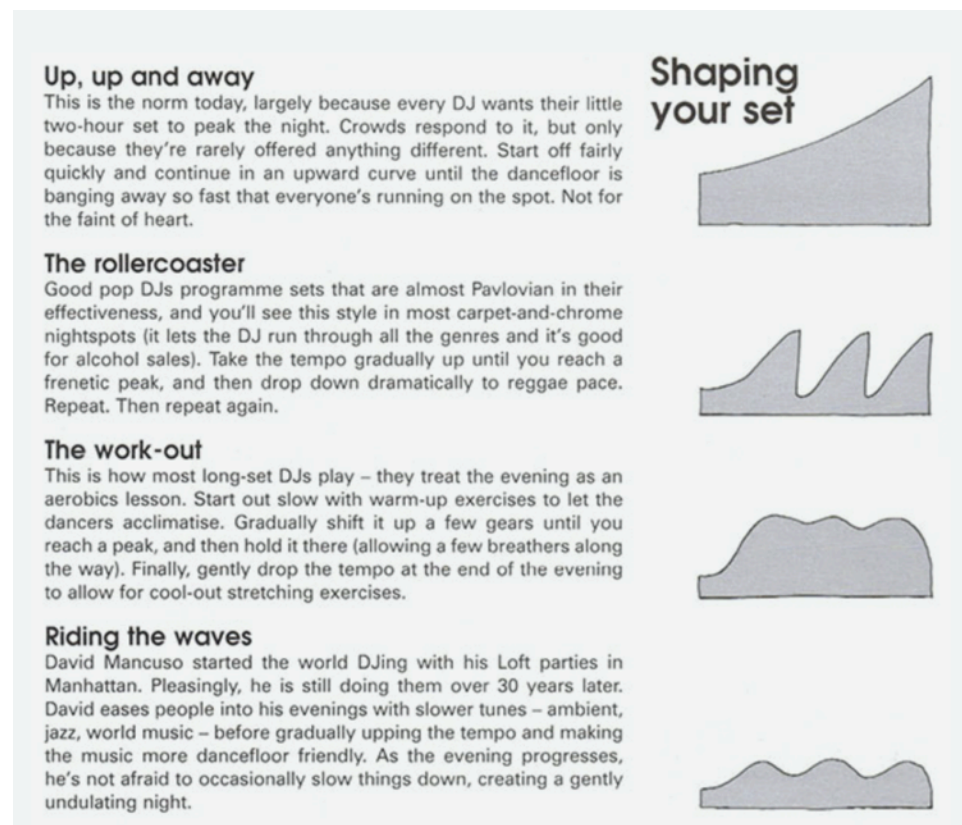


Figure 52 Four archetypical shape of DJ sets according to Broughton, & Brewster (2003, 135)

One way of creating similar curves for the DJ sets would be to attempt to quantify energy by applying machine learning techniques as proposed in Chapter 5. However, the concept of energy is also an overarching concept and not very music analytically precise, and I do not have access to a dataset of annotated energy ratings of EDM. MIRtoolbox

nevertheless enables the more detailed investigation of what musical aspects are "turned up and down" during the sets.

My first point of departure is that the audio files of the DJ sets were divided into 10-minute sections in order to be able to retrieve features from them. In the following, Section 1 denotes minute 0-10 in the audio file. Section 2, minute 10-20, etc. I only included the audio files that last between 50.31 minutes and 70 minutes. First, I chose to apply the PCA to initially explore whether there musical characteristics that adhere to the beginning, the middle or the end of a set, and whether these differ within subgenres. Are there any clear formulas that adhere to a genre?

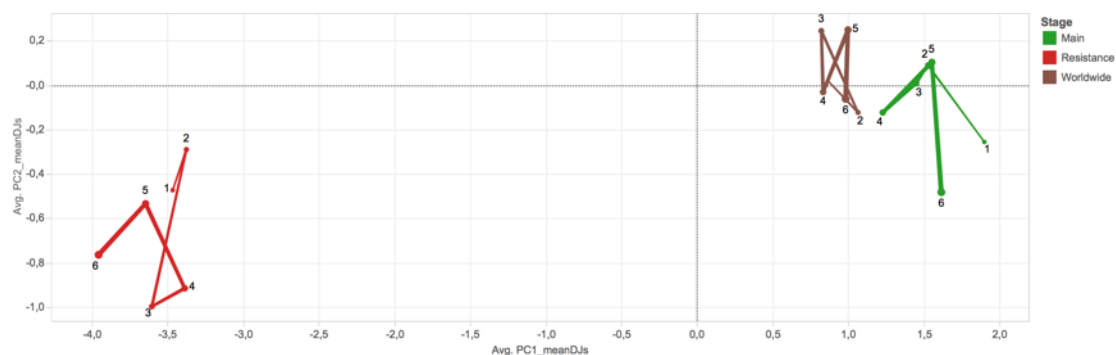


Figure 53 A plot of the average "routes" on three different stages. The labelled numbers indicate which section of the set. The values are average PC1 and PC2 values for all sets from the respective stages, in the respective section. There are 15 sets from Resistance stage, 19 from Main, and 8 from Worldwide. Appendix 4 contains a plot of PC3 and PC4.

I use this PCA plot exploratively, to map my data initially. The plot could seem to indicate a slight tendency that 1 and 6 are closer placed, which would insinuate a general tendency that the ending on the average is closer to the beginning than the parts in between. This would imply that music goes from home, diverges in some direction, returning closer to home than it was in the middle. However, this tendency should be taken with a pinch of salt, again because there are many layers of reduction that distort the picture: Firstly, the plot only accounts for about a quarter of the variance in the dataset. Secondly, because the values only are averages for many sets, if one set goes in one direction, and another in the opposite direction, these will average each other out. But again, I can diminish the amount of reduction by looking more closely at the data.

The first problem can be nuanced by statistics:

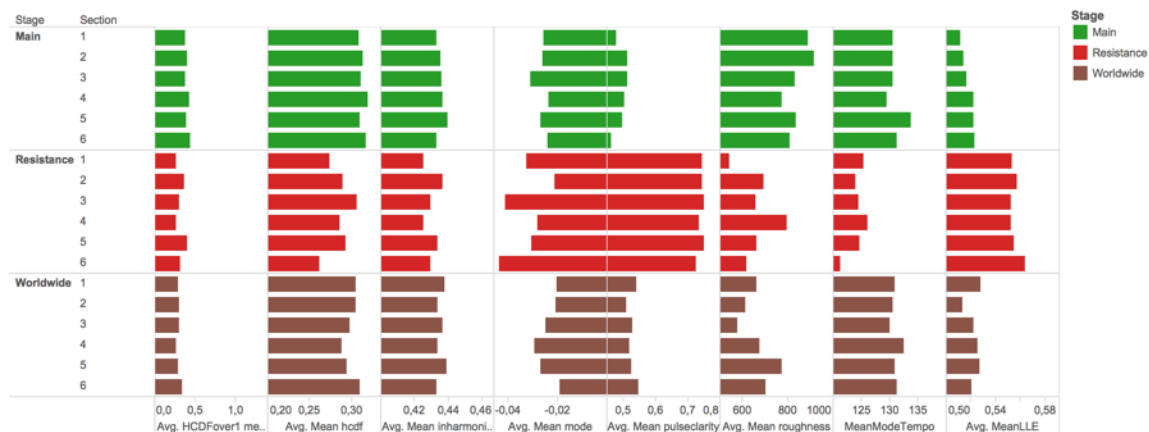


Figure 54 Average features arranged by sections and stage.

Zooming in on the Main stage, I especially note a tendency that Roughness, Zero-cross and Pulseclarity fall towards the ending of these sets. Rhythmically this implies less staccato rhythms, induced by longer, drawn out sounds more prominent in the mix. Harmonically, the music becomes less dissonant, less noisy and has more major chords towards the end. In view of the larger course of the set, these qualities could indicate a general suspense-release pattern; from tight, hard, dissonant to more loose rhythmically and more harmonious. This pattern can be found rather clearly in for example Axwell & Ingrosso, Blasterjaxx, Fedde Le Grande or Zed's Dead. Steve Angello is another example, though he raises the tempo towards the ending, thereby tipping the release with an energy outburst.¹⁵⁸ This tendency backs up¹⁵⁹ an assumption, I concocted from listening through the sets, before I analyzed the data. And the data can now help me back up this assumption.

But the data help me nuance this initial generalization. For example through plots like Figure 55. The plot displays that within the average shape are many variations that do not fit into the general pattern delineated above. Where I thought Fedde le Grand (light brown line) as the prototypical shape of a set, the reality is more diverse: 7 of the 19 sets decrease on both Pulseclarity and Roughness, when comparing section 1 to 6. But 4 increase both values, which indicates a rise in tension. Thus the data helps me nuance and quantify the general tendency.

¹⁵⁸ The data helped me identify these DJs by investigating the worksheet "Statistics_Sections_Stages_acts" in Tableau, 480DJSubsets.

¹⁵⁹ Though not tested for statistical significance.

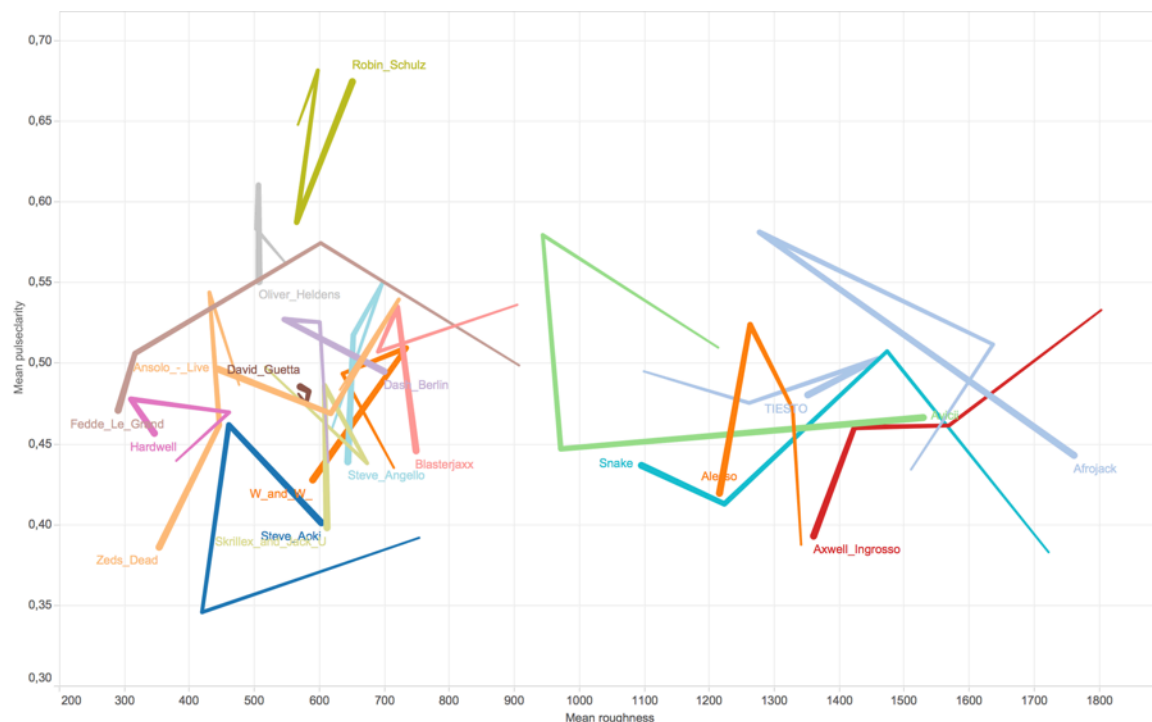
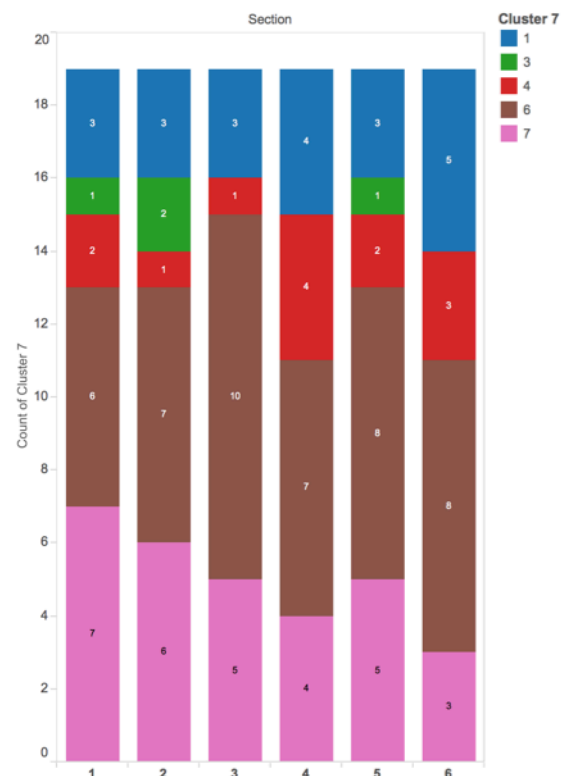


Figure 55 Mean Roughness and mean Pulseclarity for 19 DJ sets from Main Stage. Due to visual clarity, I only included section 1,3,4 and 6 in the plot. Line thickness indicates the course; thin is section 1, the beginning. Thick is section 6, the ending.

I could also choose to apply a cluster analysis technique to automatically assign a label to each section according to its features. This would provide me with another way of investigating the question about journey. For this analysis, I applied an unsupervised 7-means clustering technique to label all sections a cluster from 1-7. The cluster plot in Figure 56 demonstrates that 7 sets begin in cluster 7, but only three ends in this cluster. However, Cluster 1 and 6 become more prominent towards the end.

Figure 56. 19 sets from the Main stage mapped according to section and cluster. The numbers in the bar represent the number of sets that adhere to the cluster and section.



Musically, it is necessary to compare this knowledge to Figure 57 below, because the cluster numeration certainly does not speaks for itself.

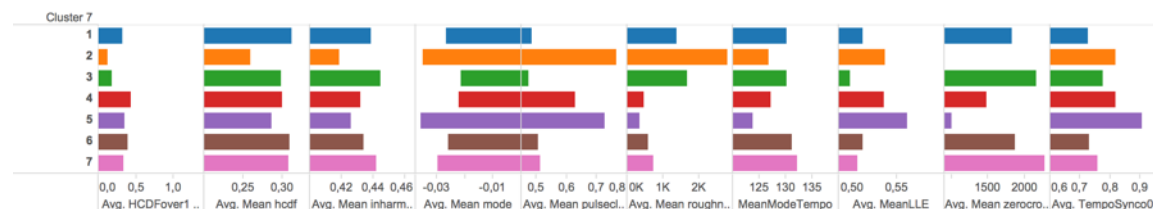


Figure 57 Statistics for each cluster.

When I compare the two Figures above, a tendency that over half of the sets that begin in the more distorted, noisy, fast cluster, but ends in other becomes prominent. On the contrary, no sets switch to cluster 7 in the final sections, unless they have begun there. On the contrary, the music that the DJs at Main Stage play is more often assigned cluster 1 and 6 in the end than in the beginning. Both these clusters indicate low Pulseclarity, and LLE, but they differ in roughness.

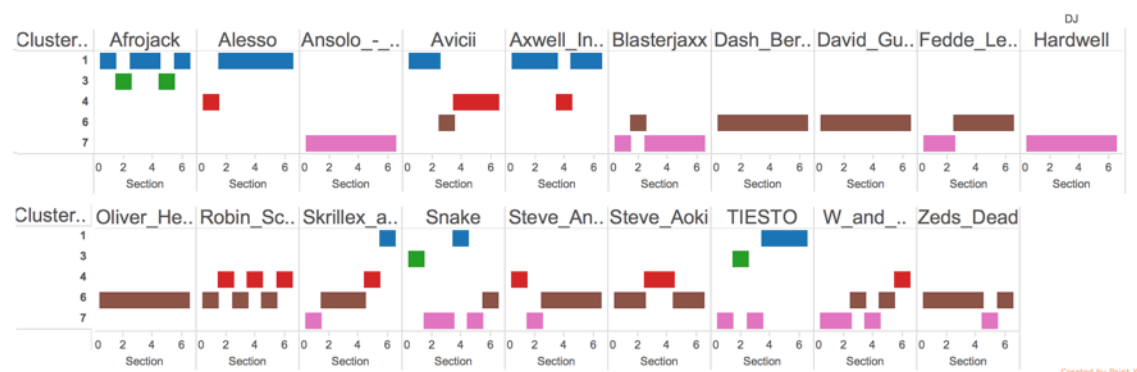


Figure 58. 19 sets from the Main stage mapped according to section and cluster.

My findings above indicate that many sets at UMF do not follow Broughton and Brewster's archetypal patterns for shaping the sets. At least when it comes to musical parameters that traditionally are associated with energy; high tempo, distortion, and dissonance. A plausible explanation could be that DJs approach short sets with another type of energy curve. In a way that is comparable to how DJ Frankie Knuckles' express his approach to short sets: "I have to give it my all from the very first record ... [there's] no build-up, no pacing" (quoted from Fikentscher 2013, 132 - squared brackets in the original). The data suggests that some of the world's most famous DJs often apply this strategy for sets lasting only one hour.

7.5 Step 5 - Exploring Compositional Traits: Analysis on the Meso-level

7.5.1 Concerns about averages and standard deviations

Though the method of averaging features over 10 minutes sections can perform useful calculations, it also has its limitations. The approach has proven successful for mapping different DJs' styles, and for learning about overall textural qualities of the music. However, the music analytical reflections that it entails are limited most of all due to the averageness; because it largely ignores temporal developments that take place on other temporal levels than the very small and very large.

When programming dance music, Fikentscher, both a researcher and a DJ, explains that he considers whether the next record will “add variety or monotony to the mix?” (2002, 95). If a DJ's ideal is tipping towards monotony, the worth of averages increases. However, if musical energy is created through contrasting quiet and noisy passages, average values will fall somewhere in the middle, and therefore they will not convey the range of modes of expressions. These statistical limitations affect the strength of the epistemological value. But they also affect the practical value of ACA methods, because there might be a lot of hidden information in the data that is not revealed this way. Extreme cases can, for example, become hidden in the midst of the measurements.

Within other fields, standard deviation would be a statistical solution to reduce some of this problem. But calculating standard deviations from MIRtoolbox' calculation makes it hard to explain whether the standard deviation accounts for fluctuations at the micro second level or at time units that can be comprehended humanly. Or put in another way; whether the standard deviation concern the musical texture¹⁶⁰ or the musical form.

¹⁶⁰ An example of standard deviations calculated at the textural level is found in George Tzanetakis, 2014, retrieved November 26, 2016. Tzanetakis compares the spectral centroid of passages in Beatles and Debussy. Beatles has high standard deviation because of its rhythmical elements, especially the bass, which causes the centroid to alternate between high and low values. In contrast, the drawn out sounds within Debussy makes its centroid rather constant.

7.5.2 Tempo and degree of change

To overcome this, I divided each DJ set into segments of 10 seconds and calculated 14 features¹⁶¹ for each segment.¹⁶² Hereafter, I applied both a 21-means clustering analysis and a PCA¹⁶³. From this, I calculated two new measures: *Cluster Length* is calculated by counting the number of segments in a row that are assigned the same cluster, multiplied by 10 since each segment lasts 10 seconds. *PC-distance* denotes the feature difference between a segment and the previous segment. PC-distance for a segment is calculated by subtracting the previous PC1-PC5 value from the current segment's PC1-PC5, with each PC weighted according to how much variance it describes.

Both PC-distance and Cluster Length are coarse measures that serve to quantify musical variance. They are created to assist answering questions about monotony or variation. PC-Distance is a measure for how much the music varies from segment to segment, and is a way of quantifying the *degree of change*. Cluster Length is a measure of how long time a certain sound idiom lasts. A corresponding music analytical explanation to the clustering process could be that the music automatically is grouped into 21 clusters containing different sound textures found in the corpus. The measure Cluster Length denotes for how long time a cluster sounds and is a way to quantify the *tempo of change*.

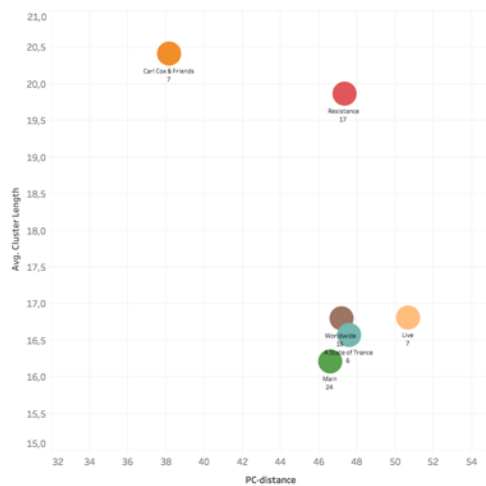


Figure 59 Average Cluster Length and PC-distance of each stage. Numbers represent number of sets taken into account.

161 Mean Zerocross, Rolloff, Brightness, Roughness, Irregularity, Flux, Eventdensity, LLE, Metroid, Pulseclarity, Mode, Inharmonicity; Mode Tempo and Metroid rounded to nearest integer, and Metroid/Tempo.

162 The script: getAllIntervalStats.m. The dataset: 480 DJ subsets, Sheet: DJ sets intervals.

163 Correlation matrix is in the sheet "Correlation Matrixes" of the dataset.

Figure 59 displays the average Cluster Length and PC-distance at six different stages at UMF. The plot seems very representative of the sound of the scenes: The dominating ideal at Carl Cox & Friends and Resistance is monotony: The tracks are merged almost unnoticeable out and in of each other. The music consists of repeated loops, and energy is built up rather continuously and slowly evolving. The remaining four stages displayed represent more a strategy of variation; they contain many contrasting passages, more discontinuous outbursts of energy and shifts in intensity.

But what do these numbers represent more precisely? Firstly, they are created from complicated processes that involve a lot of black boxing. Hence, there is no simple explanation for them, and it is rather difficult to deploy a precise music analytical language to describe them. Secondly, they are intrinsically tied to the dataset, which in this case implies that the forming of clusters and calculation of PC-scores are entirely dependent on the corpus and the 14 features calculated.

Music analytically, this corpus restraint entails that it makes no sense to compare these measurements to another corpus'. These are measures that express the tempo and degree of change. However, they cannot be standardized, only compared with other figures in the dataset. The feature restraint is determining for how I can interpret the clusters and PCs. For example, what constitutes a cluster and how does it sound like? Since the features in this dataset mostly are measuring textural aspects of the music, these are also what clusters and distances are dependent upon. I found a very typical example in Carl Cox 27March at Time 1500 and 1520 (minute 24.50-25.00 and 25.10-25.20). At that point, the music consists of a 1 bar repeated chord ostinato with effects and a bass drum playing four-on-the-floor. At 24.50 the bass drum is removed, while the ostinato continues to play, and the filters of the effects are altered slightly. The bass drum returns at 25.07. These two Times are assigned cluster changes and high PC-distance values. This is due to the effect the bass drum has on the features: At 24.50 the bass drum's pitch information no longer blurs the distinctness of the tonality, and its removal entails that MIRtoolbox measures a rather unambiguous minor chord. Furthermore, the lack of low frequencies entails a rise in Brightness, while Metroid and Pulseclarity decrease because the pulse becomes less distinct. These features determine a large part of the variation in the dataset, and consequently, large changes in them entail high PC-distance and cluster change.



Figure 60 The development of some features and throughout Carl Cox' set, March 27. Colors display the different clusters. X-axis represents time.

On the contrary, a repeated melody that profoundly changes tonality but maintains instrumentation will most likely not infer cluster change nor high PC-distance. This is so because it does not affect the features in dataset profoundly since tonal aspects are mostly ignored in the dataset.

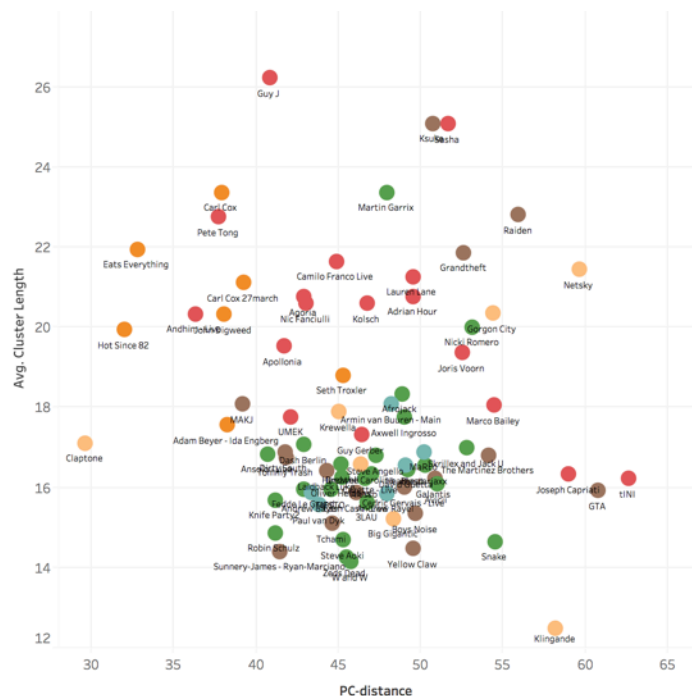


Figure 61 Average PC-distances and Cluster Length for all DJ sets.

The plotting of individual DJ sets suggests deeper insight in how to understand the numbers. (

Figure 61). The variation especially within Main Stage, A State of Trance and Carl Cox & Friends is rather small. This indicates a consistency in music programming strategies that relates to subgenres. At this festival, DJs who play similar type of music also have similar tempo and degree of change. Again the plot suggests nuances and music I can investigate further. For example, tINI's and Joseph Capriati's stand out in the plot, as red dots aside from the other red dots. This indicates that the programming tactics of these artists are different from the other acts at Resistance Stage. When I listened to their sets, a more diverse approach to music programming was revealed in agreement with the figures. Their music is based on monotonous' repeated patterns, just like the other acts at Resistance Stage. But in contrast, tINI and Joseph Capriati compose these monotonous in a non-monotonous way. This implies that the music is more active at a meso-level of about 10-60 seconds as it changes state more often than other sets at Resistance.

Ksuke's set is in many regards one of the most aggressively changing sets in the corpus. Two consecutive bars are rarely identical. The music is packed with many breaks, shifts in instrumentation, changed filters, and new rhythmic patterns. However, the algorithms, averaging over 10-seconds intervals, do not capture this large amount of musical variation. Instead, the music, measured mostly at a textural, surface level, is assigned a general state of discordance, indicated by constant high Roughness, Inharmonicity and Flux values. As a consequence, the algorithms assign long sections the same cluster label, despite that they are very diverse on both rhythmical, tonal, melodic levels. The quantification of change is constrained by the data at hand. And the interpretation of the data is bound to music theoretic notions about what is musical change or similarity.

7.5.3 Musical Development

The features can also learn us about how the music develops. For that purpose, I applied the prefix Difference which denotes a feature value subtracted its preceding feature value. If the Zerocross value rises with 100 from segment 1 to segment 2, Difference Zerocross for segment 2 becomes 100. If it falls, its Difference value becomes negative.

In this dataset, there are generally more values that represent movements towards more discordance, than the opposite. One example of this is shown in the histogram of Difference Zerocross. Similar plots of other discordance features, Roughness and Inharmonicity, could be created to tell the same story.

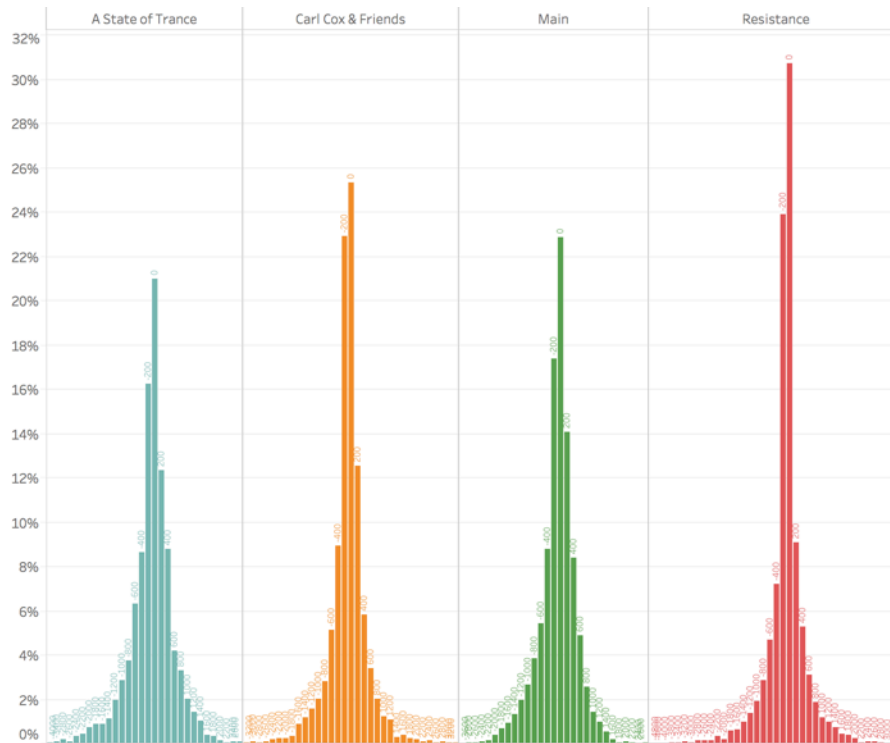


Figure 62 Histogram of Difference Zerocross values for 4 different stages. Values are binned into bins of 200. The value 0 for example represents Difference Zerocross value from 0-200 and so forth. Values are displayed as percentage of total instances within the stage.

Figure 62 indicates several points: Firstly, there are more positive than negative values, the bins from 0 to 200 and 200 to 400 are larger than those from -200 to 0 and -400 to -200. Thus the majority of sections have higher, which indicates a general movement on the 10-second level towards noisiness. Secondly, differences in stages becomes apparent. Resistance stage has the highest peaks around 0, which shows that there is less movement towards noise than for the other stages. Thirdly, all graphs are not evenly balanced: The positive values are more concentrated towards the middle, while the negative values are more spread out. This indicates a tendency that *when* the feature value drops towards less noise, it does so more profoundly.

A more clarifying way of demonstrating this last tendency is to calculate percentiles. Figure 63 displays that the median (Percentile 50) is positive for all stages. This implies

more positive than negative values, while Percentile 5 is lower than Percentile 95 is high for all stages.

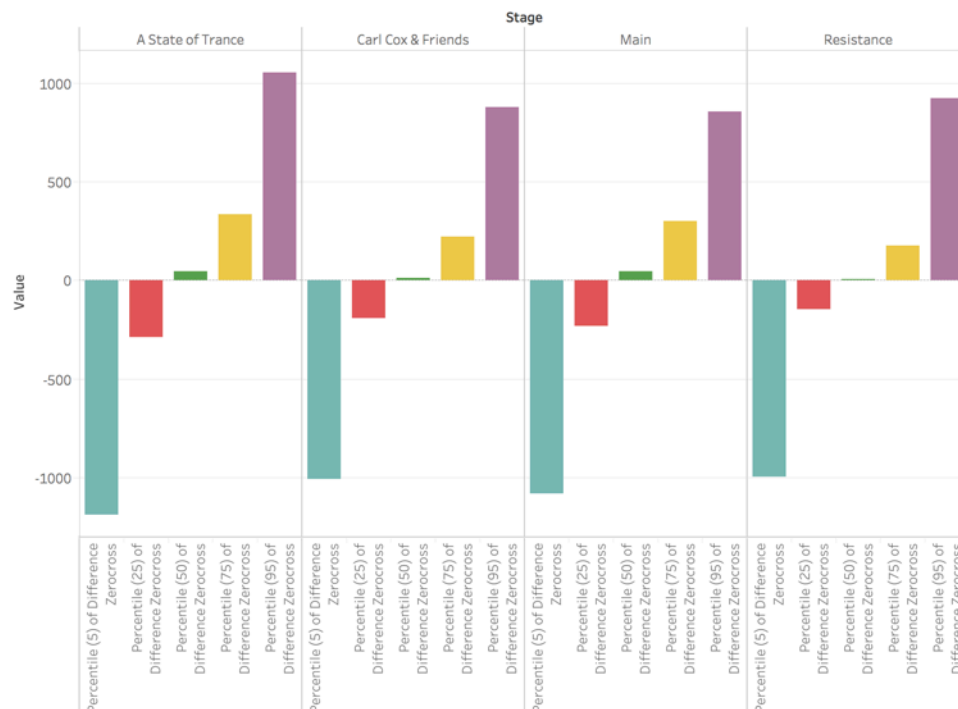


Figure 63 Percentiles of Difference Zerocross values for 4 different stages.

The graphs also expose differences within stages: A State of Trance Stage and Main Stage have both the highest median values and the largest difference between percentile 5's lowness and 95's highness. These figures are further indicators that the music develops more and faster within these stages and drops more profoundly when it drops.

Again this tendency and nuanced can be investigated closer by looking at it in another way: The PC-distance measure can be deployed for answering the question of what happens when the music changes more profoundly? In Figure 64, I have calculated Difference values for points in time where PC-distance is above 75. This plot indicates what happens to the features in those points in time where the PC-distance is above 75. Especially Main Stage stands out. The general tendency on this stage is that Zerocross, Roughness, Inharmonicity falls, while the rhythmical features Pulseclarity, Eventdensity, and Metroid also decrease in these moments. This indicates that profound changes most often imply a drop in discordance; less noise, less dissonance, fewer notes, and often also the removal of the beat, while the tempo often is not affected. Many Main Stage DJs tend to create moments where the music drops more in profoundly in intensity on many parameters before they let the music become more distorted and tense again. However, at

the DJs who plays music with many syncopations¹⁶⁴, Skrillex, Snake and Zed's Dead, do not follow this pattern. Musically, these artists represent more of high-intensity-throughout strategy, with many abrupt changes, often with sudden outbursts. This is also reflected in high tempo and degree of change values for these acts.

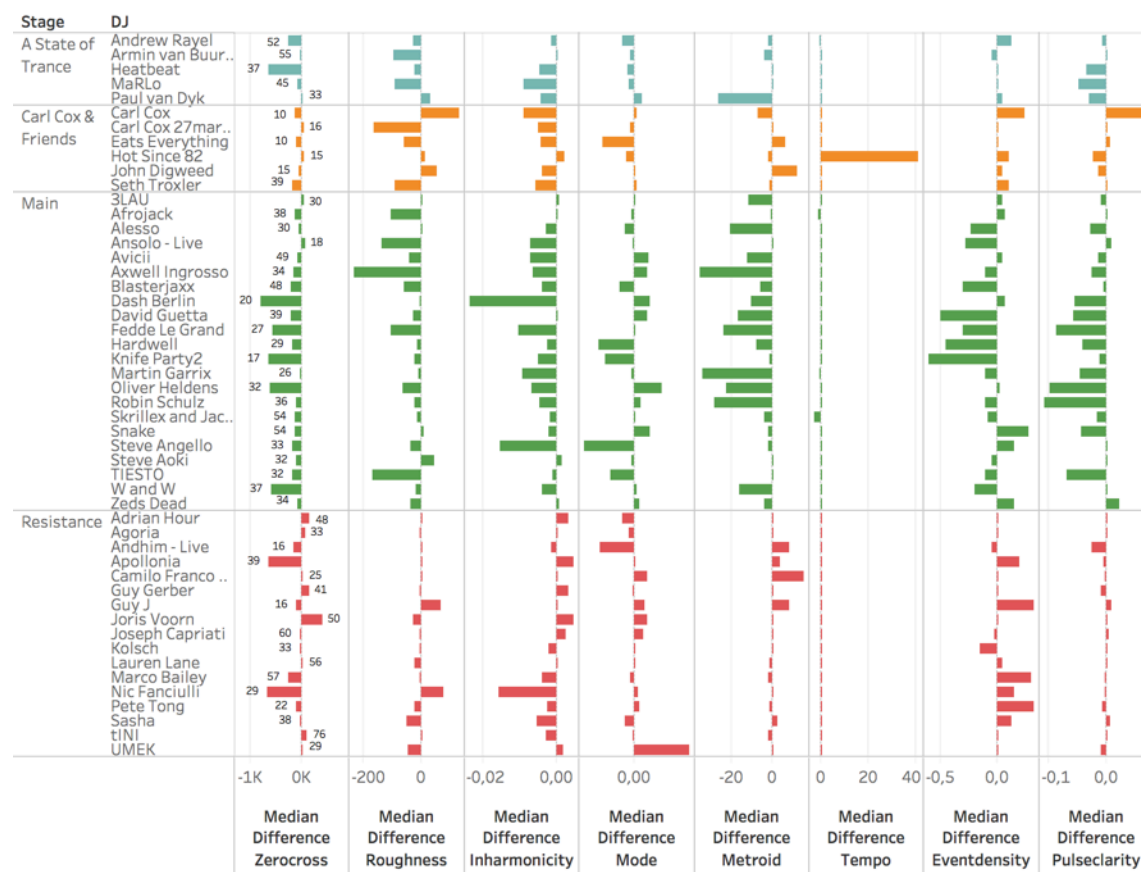


Figure 64 Median Difference values for the points in time where PC distance is above 75. The numbers besides the Zerocross values display the number of instances that are included for each DJ.

7.6 Conclusion

In this chapter, I applied some of the most common and basic low-level MIR features for a corpus analysis of 89 DJ sets. I wanted to test the methods' proficiency by applying them for an analysis of many audio files, to answer the questions I had about DJs' music programming practices. The two main questions I examined in this chapter were *how can data analysis techniques assist knowledge production?* And *what does it tell us about the music?* These

¹⁶⁴ Measured in low TempoSynco values.

two questions are interleaved because the answer for each of these questions is dependent on knowledge about the other: Data analysis techniques' usefulness is dependent on the question we can pose with them, and whether they can assist us knowing more about the music. While the methods constrain what can be measured and therefore also the type of knowledge that can be derived.

7.6.1 How can data analysis techniques assist knowledge production? – Part I

Throughout this chapter, I have sought to carefully elucidate all my steps and considerations in a music analytic process. My approach was to deploy the methods heuristically. ACA methods can create quantitative data. These data can be explored through statistics and visualization. And this can enable new ways that we can grasp a large corpus. In short, I sought to demonstrate and investigate the tools' proficiency for generalizing, nuancing, exploring, discovering, and assisting and informing listening.

The corpus I analyzed was messy. The audio and the recordings had varying quality¹⁶⁵. It contains a lot of interviews,¹⁶⁶ presenters' voices, DJs screaming, volume turned up and down, jingles,¹⁶⁷ noise from the audience,¹⁶⁸ malfunctioning sound systems¹⁶⁹, etc. But for my purpose, these errors became subordinate and somewhat unimportant in the larger picture, which interested me. This is one of the clear benefits of large amounts of data, statistics and computers in combination. We can live with messy data, amongst other also because it is possible to apply data analysis techniques to identify some of the errors.

Compared to the cases investigated in the previous chapters, both control and interpretational value were enhanced by the features being more low-level and detailed measuring at the very detailed level, and because they were accompanied by a manual (Lartillot 2014). The improved transparency certainly delimited black boxing and made the basis for understanding what is actually measured a lot better. But the element of black boxing is not entirely eliminated. The measurements are still products of very complex processes and it requires listening at a very fine-grained level, both regarding the timbral and dynamic level, for me to grasp the connection between measurements and music: It requires listening for minuscule dynamic variations to understand the rhythmical features.

¹⁶⁵ E.g. a heavily distortion in Guy J and Sasha

¹⁶⁶ E.g. in Seth Troxler

¹⁶⁷ E.g. in Armin van Buuren

¹⁶⁸ E.g. Martin Garrix

¹⁶⁹ E.g. in 3LAU, where the sound gets cut of.

And to grasp both the timbral and tonal features, you should involve your attention to the full spectrum of sound, and for example include pitch information from percussion.

Even though I was aware where to direct my attention, and even though the manual made me more conscious of the methods, it was not always feasible to establish a sufficiently useful link was not feasible in every case. Nic Fanciulli's sudden large peaks in Roughness in minute 3.30-5.00 is one example for which I have no plausible explanation.

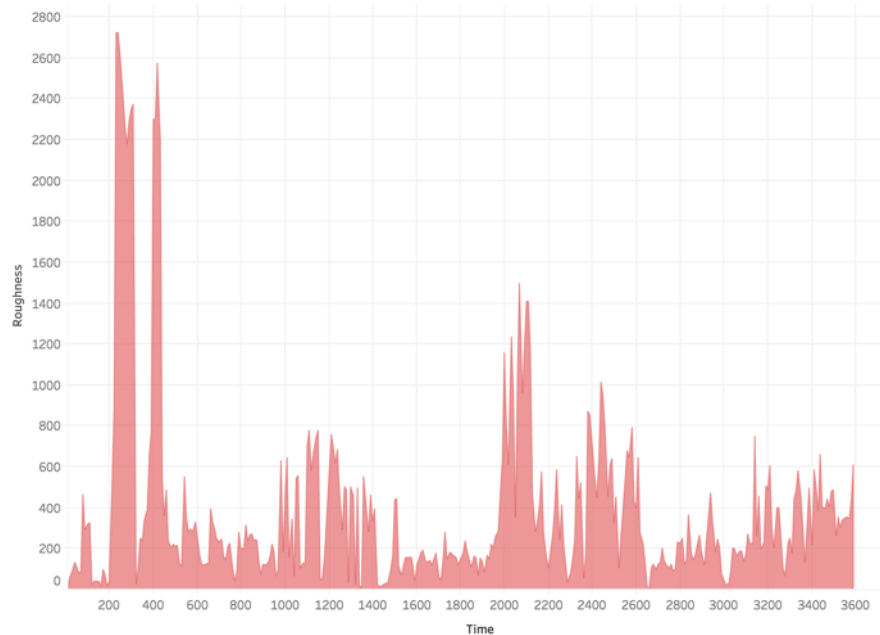


Figure 65 Roughness over time in Nic Fanciulli.

Statistics are constrained by what is measured. These methods that I applied were not only restricted to certain ways of listening to sound; they were also focusing on measuring music on its textural level. A lot of important musical information is not taken into account this way. Basic rhythm analytical concepts and important rhythmical markers, such as accentuations, subdivisions, or groove were not included particularly well in this analysis. However, the very basic MATLAB-setup enables a significant amount of improvement and control for the analyst, but this requires some programming skills. The measure TempoSynco was an example of modeling an estimator of the amount of syncopation, a measure I found useful for better my understanding of the corpus.

7.6.2 What did we learn about the music?

Despite these concerns, ACA methods proved useful for my task of assisting me in comprehending 89 DJ sets. ACA methods provided a useful means of looking at these many hours of music through the lens of data. By applying just 17 features, I could capture overall tendencies in stylistic differences and compositional traits within my corpus. These features enabled me to discern the music played at UMF from other types of music. EDM has a strong emphasis on the beat, it changes chords less often than my reference corpus, and the tempo is rather uniform across the different acts. This may not come as a big surprise to those who know about the music, but I would like to repeat my argument from Section 6.6, namely that this demonstrates that the tools work as intended, and they therefore can be applied for exploring corpora of music you do not know.

By looking at the data, I could distinguish different EDM subgenres from each other. One subgenre that stood out within the dataset was the music played at Resistance Stage. In comparison with the music at Main Stage, Resistance Stage played music that was 5-10 BPM slower but had more accentuation and space in between the beats. The differences between stages did not only concern the music's textural level, but I also found compositional differences both through listening and in the data: The DJs of Resistance Stage composed their sets more monotonous, developing more slowly. Main Stage DJs, in comparison, were changing musical states more often, and their periods of building up - breaking down were briefer than at Resistance Stage and more constrative. The data also indicated a slight tendency towards a suspense-release pattern over the course of the one-hour sets at Main Stage. Thereby the data suggested me that for DJ sets lasting about an hour, none of Brewster & Broughton's archetypical shape of sets fit. Rather I would argue that there is a tendency that many sets follow a pattern that very roughly looks like this drawing:



Figure 66 Prototype of a DJ set on Main Stage.

However, this also depends also on how you consider musical energy, and how it manifests itself in the music.

7.6.3 How can data analysis techniques assist knowledge production? - Part 2

Thus, I could use the data to back me up in what I already had experienced through listening, and what I thought I knew beforehand. I could also use the data for materializing and manifesting this knowledge in more concrete forms. Equally importantly, the data could also assist me nuance what I, with my limited mental capacity, thought I knew, which was mostly based from rough generalizing and mental classifications. I assumed, for example, that there was a general tendency on Main Stage to start up more energetic and tense and end up in a state of release. But the data indicated that this is not always the case, and provided me with estimations of how often this is not the case.

Consequently, I applied ACA methods as a tool that, with the words of Ian Foster, can “enhance perception”. Foster writes that “[i]nformation technology can [...] enhance our abilities to make sense of information, for example by allowing exploration via visual metaphors” (2011, 19). This describes well what I did and what I gained from doing it. However, this enhancement or expansion has another form than musicologists traditionally are used to. It takes other skills than previous to bring it into play: Data, statistics, and visualizations are the medium. And it takes the language of uncertainties, indications, suppositions, suggestions, coarse generalizations, reservations, doubts, etc. to arrive at qualitative inferences. I also see signs that we will have to be able to live with black boxing if we want to find trends in large corpora of audio files. But I at the same time also see indications that it *is* possible and beneficial to deploy ACA methods in practices despite black boxing. The methods will still enable us to enhance our perception.

CHAPTER 8

Conclusion

This thesis was born out of the new potentials, and the recent trends within the humanities of applying digital techniques for the analysis of many cultural objects at once. Musicology is facing similar conditions as other humanities fields: There are large amounts of digitized audio files, and there are digital tools that can analyze these audio files. I extracted the research question out of these circumstances:

How can ACA methods be used for conducting large-scale analyses of audio files of western popular music for musicological purposes?

To answer this question, I chose to seek to establish a connection between two separate fields: The field of Music Information Retrieval develops the ACA methods and is based in a computational tradition. While musicology, which is the field I investigated these methods' usefulness for, is primarily affiliated with a humanities tradition. Hitherto, these two disciplines have existed largely independently with only sparse communication between them. This dissertation pursued bridging between not only from one discipline to another but also from one research culture to another. In order to approach this gap, I amongst others turned to the broader fields of Digital Humanities for general theories on how to incorporate digital methods for large-scale analyses of cultural objects. And I also consulted digital humanities' studies and experiments for inspiration.

8.1 What to win?

My first sub-questions was:

What new potentials arise for musicologists when applying ACA methods?

My point-of-departure was the new potentials the tools could enable, and which I saw other humanities' fields were beginning to exploit, but not really musicology. Computers allow us to perform tasks a lot quicker and much more persistently than humans. This

advantage enables us to calculate aspects of more music than we can ever listen to. As Foster explains, computers can help us “enhance perception” (2011, 19). They enable us ways of looking at our objects that would not be possible (or would require a lot of practical work) otherwise.

ACA features comprise of quantitative measurements of audio files. When we apply ACA methods for analyzing audio files, we, therefore, enter the realm of data analysis and statistics. This brings along some advantages. We can use statistics to create compact descriptions of our objects, to create overviews of large entities that are hard to grasp otherwise. Generalizing is a means of grasping more objects at once. The study by Mauch et al. (2015), which I investigated in Chapter 6, provides a good example of this. They applied a range of statistical reduction techniques to squeeze the timbral and tonal complexities of five decades of popular music down to 16 topics containing musical archetypes. This reduction enabled them to grasp, identify and visualize overall tendencies in their corpus.

But statistics are not only useful for generalizations. In Chapter 7, I applied data visualization techniques for shuttling between an overview of the overall tendencies for my corpus of 89 DJ sets and statistics for grasping the individual DJ sets. The data allowed me to nuance and find variation within the overall categories. This alternation was made possible from the quickness of inquiry that software visualization tools enable.

Music is complex, and this complexity is not easily to account for by calculating a few features. Therefore, I especially found it useful to exploit the power of dimensional scaling techniques that can reduce high-dimensional datasets down to more manageable sizes. I found the principal components analysis (PCA) very helpful for exploring and finding overall patterns in these large datasets. The PCA does not require a hypothesis beforehand. In my study of DJ sets, I applied PCA for mapping my objects. Through this operation, I identified that DJs that played at the same stage also had feature similarities. This indicated that there was genre characteristics that adhere to the stage the DJs played, and the 17 ACA features, I had ascribed each DJ set, were enough to be able to discern different sub-genres from each other. The PCA can also be applied for analyzing how much the variables mutually correlate. This is a significant advantage especially in the early stages of using ACA features for music analysis, where there is a lack of experience conducting music analysis with the features.

The last benefit I have covered, though mostly theoretically, is that we can use data-driven methods for investigating questions that concern importance, questions that ask what

matters in the music? By connecting an annotated dataset with the corresponding audio files' features, the computer can calculate how the features correlate with the annotations. It enables us to examine from a data point of view questions such as are there common musical traits that make critics like it? Or in music from Denmark? We can apply data-driven methods for analytical purposes, and this approach can comprise a way to re-actualize and re-investigate old discussions from the perspective of data. In my case study 3, I apply a data-driven approach to find stylistic differences between sub-genres.

8.2 What to learn?

To release these potentials, we have to look at our objects through the “lens of data” (Aiden & Michel 2013). But the clarity of the view is constraint by the measurements and the data view can never be better than what is measured. Consequently, my second sub-question arise in continuation of the first:

What can musicologists learn from large-scale analyses?

At the data level, one of the prospects for applying ACA is that it enables the opportunity to work with data that is closer related to the sounding end product than previously. All the information about the sound is potentially in the audio file. In comparison, the traditional object of music analysis, scores embody primarily formalized rhythmical and tonal properties of the music. Fundamentally, ACA features are derived from spectral and dynamic information, and this basis influences all other feature calculation (Müller 2011). Precise tonal information can therefore, for example, be difficult to obtain because there is a lot of information in the signal that adds noise, such as overtones or pitch information in the percussion. This information has traditionally not been regarded as adhering to the notes played. In general, though many ACA methods seek to simulate traditional music analytical procedures, none of them are perfect, and in general their accuracy is estimated about 70-85%. Therefore, we can never be sure, whether the computational measurements correspond with how we as humans perceive the music, whether the individual calculations agree with traditional music analytical notions of the music.

However, a positive side of ACA methods is their flexibility. Features and their precision can be improved. We can also modify them, so they fit with the questions we have. For my case study, I used the tempo miscalculations for estimating the amount of syncopation, which was a useful metric for discerning between subgenres. We can also create

compounds of features to estimate and simulate more intuitively understandable aspects about the music. Echo Nest's features that seek to measure, for instance, danceability, energy, and acousticness, are examples of this approach.

However, at the epistemological level complications and limitations arise. My investigation of Echo Nest's machine-learned features demonstrated a general tendency found throughout my case studies: Though the features are objective in the sense that they are reproducible, they are very complicated to account for qualitatively, and this can impede the translation from feature to music analysis. In the case of Echo Nest, features that at the apparent level seem intuitively easy understandable are, in fact, very complicated to convert to qualitative aspects, when scrutinizing them closer. In my analyses, I have found it useful to apply *black boxing* as a central concept that pervaded many of my findings. It describes the act of putting something into the machine, but not being entirely sure of what is coming out, and why it is coming out with this result.

The lack of transparency is a primary reason why all my case studies are full of maybes, indications and uncertainties. At times, I even had to refrain from attempting to provide a proper music analytic answer, because the epistemological value of the features was too fuzzy and opaque. The Mauch et al. study, which I covered in Chapter 6, provides an excellent example: The many layers of statistical reduction they performed to make the data manageable rendered the link from data analysis to music analysis blurry and opaque.

These interpretational challenges affect when we scale up the analysis to encompass large datasets. On the one hand, when analyzing large datasets, we can live with insecurities and uncertainties, as Mayer-Schönberg & Cukier (2013) advocate. On the other hand, we should be alert for systematic biases. And as I have advocated, these biases can exist both on the level of the corpus and on the level of methods. My investigation of Mauch et al. provided examples of both. When attempting to measure music similarity, their methods bias the results in certain directions. I found that their study revealed more about the music technological developments and the US mainstream's acceptance of these, than it did about the revolutions in the music itself (whatever that means), as the researchers claimed. In any case, a general problem with opaque algorithms is that it becomes difficult to grasp whether and where systematic biases occur. Hence, they require interpretation that takes both music and statistics into account.

8.3 How to incorporate?

My third sub-question was:

How can musicologists incorporate ACA techniques into their practices?

It requires some understanding on different levels to incorporate the tools into practice. At the music analytic level, musicologists will, especially in comparison with analysis of scores, have to activate new ways of listening to music to understand the feature calculations better. On the one hand, the methods are flexible, and it is hard to put a definitive stamp on what aspects of the music they measure. However, in my analyses, I found a tendency that we will have to pay more attention to timbral aspects, listen for the full spectrum of sound. Many features are very affected by mastering, and factors such as mixing and EQ can affect the feature calculation profoundly. In the features I investigated, I also found a tendency that many current features measure aspects of the music's textural qualities, but largely ignores the music's progression

A general learning, though, is that the opacity of the methods impede the link between music and how ACA methods datafies it. It will require practice, and the establishing of new best practices, to create proper links. I have advocated that we deploy methodology-informed listening, a mode of listening where the features are sought understood by combining listening with knowledge on the algorithms. For the higher-level features, this implies that we should be aware that “many MIR systems that appear as if they are “listening” to the music when they are actually just exploiting confounded characteristics in a test dataset”, as Sturm and Collins have explained (2014, 1). The Echo Nest measure Danceability is an example of a feature that consists of compounds of other features. For music analytically grasping Danceability values, we have to find out, through listening or investigating the algorithms, what these components are. As a general approach, I have argued that we seek to investigate the relationship between data and musical aspects through practice, engaging and empirical investigation of the features relation to the music. For in many cases we will never be able to grasp the link entirely theoretically, since the algorithms are too hard to grasp, and even the audio format affects them. In any case, we will also have to get used to black boxing, implying that for most features, we will probably never be entirely able to understand how they translate into music aspects.

The other aspects of understanding is that what we will have to learn what is possible at all, and learn how to understand music in terms of data. How we can think music with

data. I have already provided examples of the potentials, when I answered the first sub-question. At the level of skills, it will be a process of creating connections between data practices and theory. New questions that arise are, what data can we create that can assist answering the question we have? What data is obtainable at all? Which new questions can be posed with the data we have? And what answers can we derive?

At the level of conducting humanities research, the tools do not require that musicologists dismiss their core competencies. But it depends on factors such as the question, purpose, and available software. For example, does the PCA mapping technique not necessarily require data analytic skills; it is just a tool that can map a multidimensional dataset. However, to properly understand what a feature measures, or to create new features, it will require a combination of music analytic and mathematic, statistic skills.

I have discussed that humanists are trained in, skilled at and prone to certain ways of thinking: Close rather than distant reading, nuancing rather than generalizing, complicating rather than simplifying, qualitative conclusions rather than quantitative, interpretation over verifiable knowledge, etc. But applying ACA methods do not render these core competencies useless. There are ways, as Wallmark explains, “we can use empirical methods without allowing the discipline to be colonized by empiricism” (2013). Firstly, because we will have to interpret the results, and interpretation requires “domain knowledge”, as Dalton & Thatcher (2014) amongst other advocate. Secondly, humanists tend to ask questions of *why* and *how*, while quantitative methods are capable of answering and posing questions of *what*. But this discrepancy does not render the tools useless. Rather, ACA methods can be applied for empirically informing the why and how-questions. They can become integrated as a part of the evidence-gathering process for answering other types of questions. As for example Cook and Clarke (2004) explain, there always is some amount of empirical observation in musicological studies. ACA methods can help us provide empirical information, as they can help us see things we would not be able to, due to our limited time available for listening.

--- --- ---

If I were to provide a clear answer to the question, whether **musicologists can use ACA methods for looking at many pieces of music at once**, it would be a “*Yes!*” succeeded by an “*it depends.*” For the answer to the question depends on how useful what they see is, and usefulness depends on the purpose.

In this study, I chose to uncover purposes ranging from the practical value to the epistemological value of the tools. I also sought to expose how these purposes relate to the

general level of conducting humanities research, and how they alter the music analytic level.

What happens in the meeting between ACA, many pieces of music, and musicology? The tools enable us to enhance the listening at several levels: When user-friendly software arrives, they will eventually provide a swift way of grasping more music, primarily by visual means. At the level of attention, these tools can suggest musical aspects of listening for. And due to their sheer way of measuring they suggest and invite us to listen to new levels in the music, and thereby perhaps encourage us to expand our listening to pay attention to detail in new ways. At the same time, this expansion takes another form than previously: The medium is visualization or statistics, which will become a new supplement to listening.

The answer to what skills it requires to attain this enhancement is also “it depends”: You can apply the methods for explorative inspiration, such as rough mapping without knowing anything about how data has been generated and what it represents. But if you want to know about what is measured, you have to dig into statistics and mathematics to some extent and combine it music analytic knowledge. But there is always the risk here of ending up in an endless hole-of-methodological-immersion. Everything can always be better. Everything can be examined more, and this goes also for algorithms. As well as how they relate to the music. As a result, an apparent paradox emerges ACA methods, which, in fact, are exact measurements of the music, translate dubiously into music analysis, and consequently they become qualitatively imprecise. And this is also why my interpretations of the data throughout the thesis have been full of *probably's*, *maybe's*, uncertainties, indications, etc. But this is a price we have to pay to generalize and be able to see larger patterns.

Uncertainties can always be diminished further, if we examine them closer. And the possibilities for criticizing large-scale analyses based on the methods, and finding flaws in the analyses, will probably be enhanced compared to not applying them. However, this itself is not an argument for dismissing the methods. We can apply them for pragmatic purposes, as a new source for empirical information that can be included in the investigation; for inspiration, for including and grasping more songs, posing new question, for expanding the views, and for helping us “shut our strong eye.”

Appendices

Appendix 1. Resources (in Spring 2016)

(All figures are retrieved February 2016).

User-friendly databases that contains music related information

Peachnote¹⁷⁰ Viro, 2011 is currently the closest you get to the musical analogue to the Google NGram¹⁷¹ viewer. It allows users to search for sequences of notes or chords and plots the occurrences of the particular sequence into a timeline. Similarly as the Google NGram. “The data for Music Ngram Viewer comes from around 1,600,000 OMR’ed sheets from some 160 thousand scores from the Petrucci Music Library, the Library of Congress and the Duke University Library. They contain almost 370 million notes.”¹⁷²

Digital Music Lab VIS¹⁷³ allows users to explore statistical properties of the music in four collections of music recordings in various genres: The British Library Sounds Archive,¹⁷⁴ CHARM,¹⁷⁵ and ILM¹⁷⁶. It allows you to create subsets (for example all music from a specific country, by a specific composer, or all works with andante in the title) and explore primarily tempo or tonal properties and compare different subsets. The information is derived via audio content analysis of these collections’ recordings.

Musixmatch¹⁷⁷ is both an easily searchable website that contains lyrics from more than 7 million songs in more than 50 languages. It comes with an API¹⁷⁸.

¹⁷⁰ www.peachnote.com/, I recommend you start at the “about page”: <http://www.peachnote.com/info.html>

¹⁷¹ In which you can search for occurrences of words or n-grams (a sequence of words) in the digitized books in Google Books corpus. <https://books.google.com/ngrams>

¹⁷² <http://www.peachnote.com/info.html> - retrieved February 24, 2016

¹⁷³ <http://mirg.city.ac.uk/dml-vis/>

¹⁷⁴ 30,000 from the World and traditional music collection, and 20,000 from the Classical music collection. www.sounds.bl.uk, retrieved February 2016

¹⁷⁵ 5,000 recordings of classical music, charm.rhul.ac.uk.

¹⁷⁶ 287,000 individual recordings in various genres. ilikemusic.com

¹⁷⁷ www.musixmatch.com

¹⁷⁸ <https://developer.musixmatch.com/> or <http://labrosa.ee.columbia.edu/millionsong/musixmatch>

Musicgraph¹⁷⁹ contains overviews of artists, a few acoustical features, social media stats and biography.

Chordify¹⁸⁰ that contains information on chords from popular songs.

WhoSampled¹⁸¹ holds information on what samples are used in more than 368,000 songs.

Everynoise.com is a genre map of 1387 genres that each contains more zoomed in sub maps of the most prominent artists within each genre.

Music Timeline¹⁸² is another example of genre visualization. The rise and fall of the most popular genres and subgenres is visualized as a function of time.

These are examples of some of the most user-friendly websites. But there are a many other databases that contain huge amounts of information on music not mentioned here.

DBtune¹⁸³ is a linked data initiative that links information in some of these that holds semi-structured data and provides access to their data.

Software

Lerch (2012) provides a list of software for audio analysis¹⁸⁴.

During my project, I tried the following applications for feature extraction:

Sonic Visualiser (Cannam, et al. 2010) but found that it was not able to conduct analysis of bulks of files, but in stead better suited for more detailed analysis of one or a few audio files.

JAudio¹⁸⁵ (McKay 2010), but I found that the features were too low-level and difficult to apply for music analysis, despite they could be extracted into a spreadsheet format. Perhaps JAudio is better suited for machine learning tasks such as classification of music as proposed in the introduction of (McKay 2010): “[I]t is designed to be a general-purpose

¹⁷⁹ www.musicgraph.com

¹⁸⁰ <https://chordify.net/>

¹⁸¹ www.whosampled.com

¹⁸² <https://music-timeline.appspot.com/#>

¹⁸³ <http://dbtune.org/>

¹⁸⁴ At the moment of writing, January 18, 2016, the list is updated at <http://www.audiocontentanalysis.org/software/>

¹⁸⁵ See here for description: <http://jmir.sourceforge.net/jAudio.html>. Or here for downloading the files: <http://sourceforge.net/projects/jmir/files/>

toolkit that can be applied to arbitrary types of music classification” (3).

The Echo Nest API¹⁸⁶ is not a software (and not mentioned in Lerch 2012), but capable of feature extraction. However, the features are not sufficiently fine grained for the analysis level, I acquired.

Datasets

The Echo Nest API¹⁸⁷ contains more than 1 trillion data points¹⁸⁸ of more than 36 million songs, including the Spotify catalogue. The features calculated are superficial and not constructed for musicological purposes, however they can be applied for creating initial overviews of large corpora of music. See also Chapter 5.

The Million Song Dataset¹⁸⁹ (Bertin-Mahieux, et al, 2011) is an often used and cited dataset. It contains a million songs, but not the audio files, only their metadata. The ACA metadata is derived from Echo Nest, while other kinds of metadata is in the dataset as well. For instance bibliographic metadata from MusicBrainz (Swartz 2002), lyrics from the MusiXMatch dataset,¹⁹⁰ last.fm tags, etc. The introduction of the dataset (Bertin-Mahieux, et al. 2011) does, however, not contain considerations regarding how the music was selected for the corpus which make music analytical claims derived from the whole corpus more difficult. (There is a complete list¹⁹¹ of tracks, but it is very enormous and unmanageable).

The McGill Billboard Dataset (Burgoyne, et al. 2011) contains metadata from more than 1000 songs from the Billboard chart from 1958-1991, chosen from a principle of randomness. This dataset has been manually annotated by music experts, containing chord transcriptions, information on instrumentation and Echo Nest features. The advantage of manual annotation is that MIR researcher can test their algorithms how precise they are against this corpus.

More comprehensive lists of datasets related to MIR can be found at:

<http://www.audiocontentanalysis.org/data-sets>.

http://grh.mur.at/sites/default/files/mir_datasets_0.html

¹⁸⁶ <http://developer.echonest.com/>.

¹⁸⁷ the.echonest.com

¹⁸⁸ <http://the.echonest.com/>, retrieved January 18, 2016

¹⁸⁹ <http://labrosa.ee.columbia.edu/millionsong/>

¹⁹⁰ <https://www.musixmatch.com/>

¹⁹¹ http://labrosa.ee.columbia.edu/millionsong/sites/default/files/AdditionalFiles/unique_tracks.txt

http://colinraffel.com/wiki/mir_datasets

MIR Communities

MIREX¹⁹² evaluates the methods for the various MIR tasks (see J. S. Downie 2003b, or for a report on the community's first 10 years (J. S. Downie et al. 2014).

ISMIR - MIR research activity is centered around International Society for Music Information Retrieval (ISMIR), which is the “research forum on processing, analyzing, searching, organizing and accessing music-related data.”¹⁹³

Appendix 2. Correlation Matrix for Echo Nest Features

	energy	liveness	tempo	speechiness	acousticness	mode	time_signature	valence	danceability	tempo_confidence	time_signature_confidence	key_confidence	mode_confidence
energy	1.000	.194	.384	.357	-.728	-.099	.129	.466	.251	.323	.307	-.035	-.173
liveness	.194	1.000	.066	.172	-.202	.000	.041	.144	-.048	.014	.035	-.054	-.062
tempo	.384	.066	1.000	.186	-.349	-.075	.211	.212	.267	.182	.099	-.151	-.182
speechiness	.357	.172	.186	1.000	-.265	-.110	-.017	.266	.182	.044	.057	-.059	-.139
acousticness	-.728	-.202	-.349	-.265	1.000	.003	-.191	-.450	-.238	-.184	-.169	.020	.129
mode	-.099	.000	-.075	-.110	.003	1.000	-.058	.067	-.030	-.073	-.104	.168	.167
time_signature ^a	.129	.041	.211	-.017	-.191	-.058	1.000	.129	.061	.049	.066	.064	.019
valence	.466	.144	.212	.266	-.450	.067	.129	1.000	.435	.128	.156	.039	-.020
danceability	.251	-.048	.267	.182	-.238	-.030	.061	.435	1.000	.200	.149	-.043	-.140
tempo_confidence ^a	.323	.014	.182	.044	-.184	-.073	.049	.128	.200	1.000	.912	-.169	-.363
time_signature_confidence ^a	.307	.035	.099	.057	-.169	-.104	.066	.156	.149	.912	1.000	-.171	-.364
key_confidence ^a	-.035	-.054	-.151	-.059	.020	.168	.064	.039	-.043	-.169	-.171	1.000	.824
mode_confidence ^a	-.173	-.062	-.182	-.139	.129	.167	.019	-.020	-.140	-.363	-.364	.824	1.000
Dimension	1	2	3	4	5	6	7	8	9	10	11	12	13
Eigenvalue	3.361	2.032	1.408	1.092	1.071	.965	.782	.698	.679	.450	.241	.143	.079

a. Missing values were imputed with the quantification of an extra category.

Figure 67 Correlation matrix for Echo Nest's features, Figure 17

¹⁹² http://www.music-ir.org/mirex/wiki/MIREX_HOME

¹⁹³ ismir2015.uma.es, retrieved January 18, 2016. For reflections on ISMIR's first 10 years see (J. Downie, Byrd, and Crawford 2009). For an analysis of the yearly ISMIR conference's proceedings see (Lee, Jones, and Downie 2009).

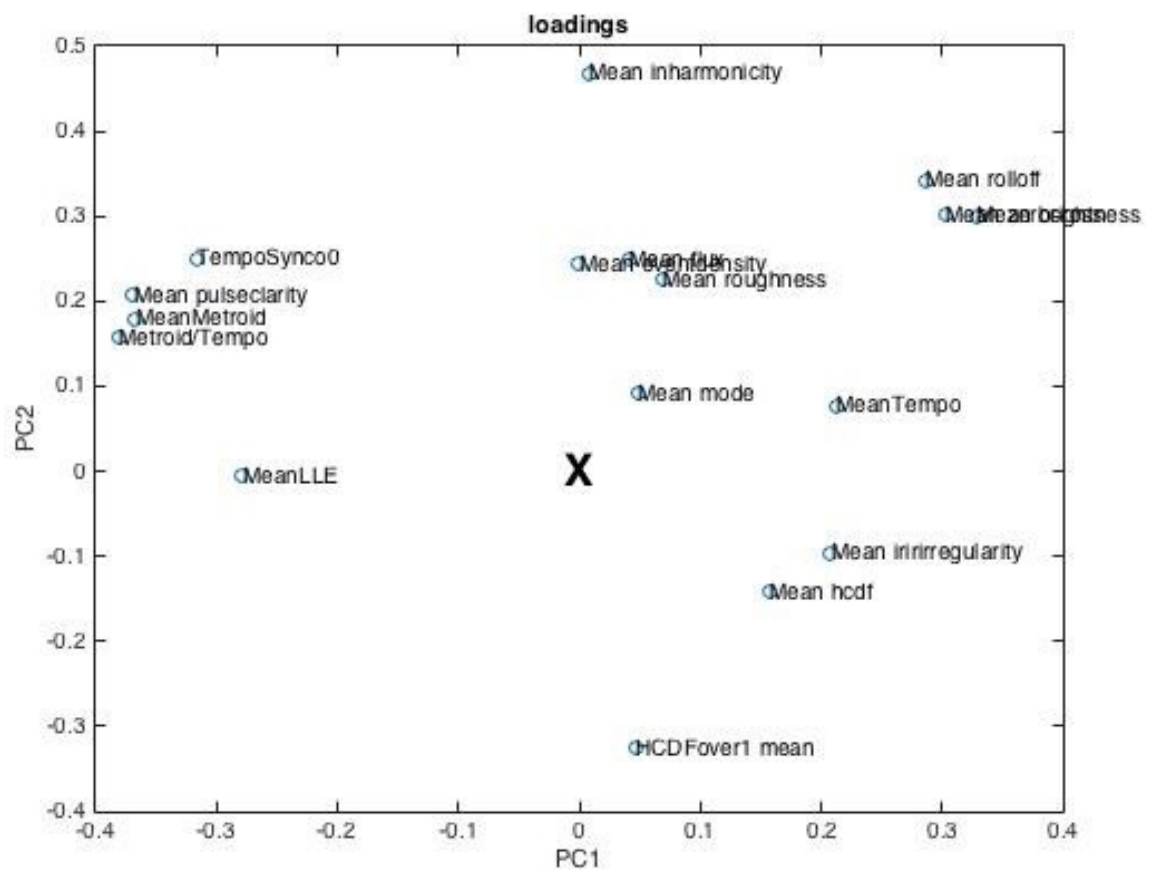


Figure 68 Loadings for Figure 51

Appendix 4. PC 3-4 for DJs and the Reference Corpus

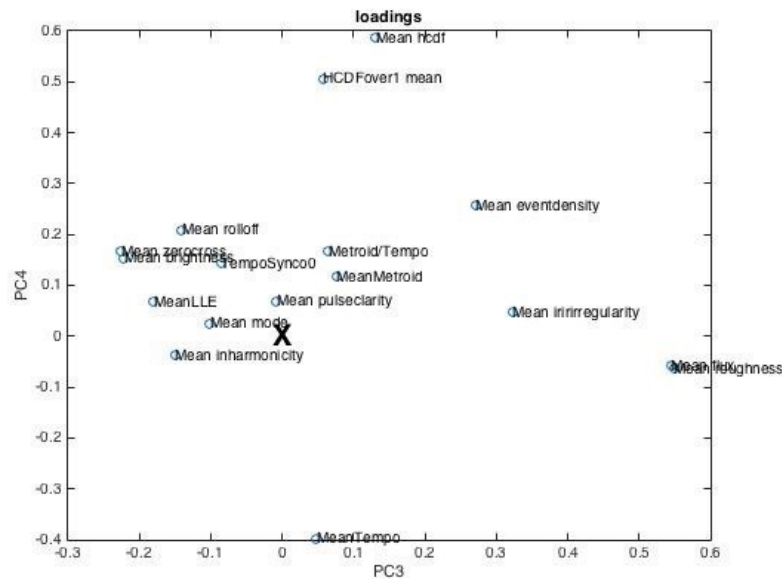


Figure 69 Loadings for PC 3-4 for DJs and the Reference Corpus

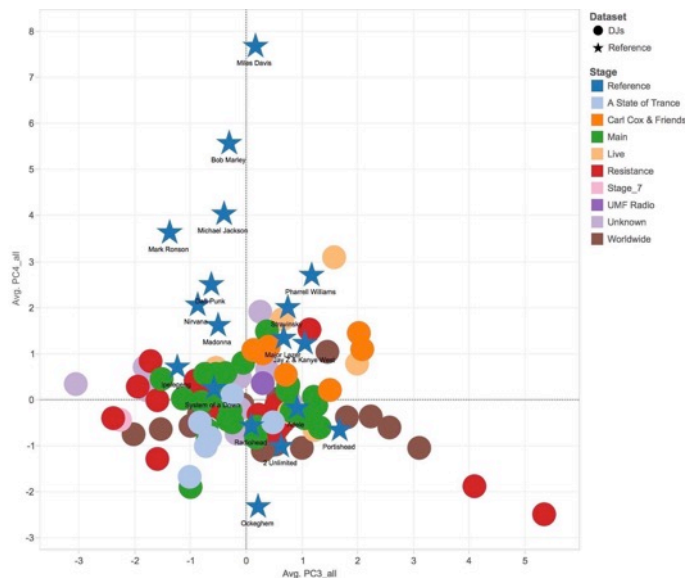


Figure 70 Scores for PC 3-4 for DJs and the Reference Corpus

PC3 and 4 helps discerning the reference corpus and the DJs further. Most of the songs in the reference corpus fall outside the groups of DJ sets.

Appendix 5. The journey of the average set

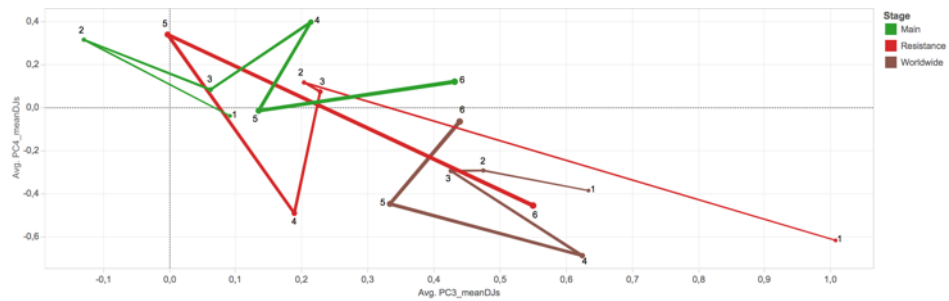


Figure 71 Average PC3 and PC4 for sections, and stages.

Table of Figures

Figure 1 Gartner's Hype Cycle of 2013.....	9
Figure 2 A growing share of digitized data is audio-visual data.....	10
Figure 3 Balen et al.'s diagram of how the born-digital metrics depend on each other.	34
Figure 4 Diagram displaying differences in end goals between Musicology, MIR and me.	42
Figure 5 The research components of my interest.	51
Figure 6 The upper-left fraction of the Everynoise.com genre map.	76
Figure 7 Three different "views" on the objects. These views ought to inform each other.....	78
Figure 8 A music analytic diagram of the flow, from music through the creation of features to understanding the music.....	81
Figure 9 Echo Nest's audio summary features for three songs..	100
Figure 10 Distribution of <i>energy</i> of the 16 different versions of <i>Sex Pistols Anarchy</i> in the U.K. in the dataset.	101
Figure 11 Tempo graphs from Echo Nest's (2013a) and Schellenberg and von Scheve's (2012, 200, figure 1) estimation of the average tempo of Western popular music.	105
Figure 12 Flowchart of machine-learned features.	108
Figure 13 Björk's songs 1990-2011 arranged according to danceability and release year of the album on which they were released.....	109
Figure 14 The development of Björk's oeuvre measured in Echo Nest features.....	111
Figure 15 Average features of Björk, compared to Radiohead.	112
Figure 16 The development of Björk's and Radiohead's features.	112
Figure 17 PCA plot of songs by Björk & Radiohead, below is a plot of the loadings.....	114
Figure 18 PCA plot with loadings of songs by 7 Danish artists.	115
Figure 19 The Evolution of topics (Mauch et al. 2015, p.4 Figure 2).....	123
Figure 20 The Evolution of styles (Mauch et al. 2015, Figure 3)	123
Figure 21 "Musical revolutions in the Billboard Hot 100" (2015, Figure 5).	124
Figure 23 Correlation between mean RMS-median RMS and lowenergy values for the reference corpus..	148
Figure 24 The RMS energy curve of Ockeghem.	153
Figure 25 The RMS energy curve of Miles Davis.	149
Figure 26 Plots of the energy curves of the first 10 minutes of a subset of the DJ sets.	150
Figure 27 LLE values for the reference corpus.....	151
Figure 29 The reference corpus' mirtempo values calculated as a function of time.	152
Figure 30 Tempo histograms of the reference corpus.....	154
Figure 31 TempoSynco measures of the reference corpus.	155
Figure 32 Median metroid values for the reference corpus.....	156
Figure 33 Mean metroid values plotted against and Mean metroid/tempo values for my test DJ subset. ...	157
Figure 34 Pulseclarity values for the reference corpus as a function of time in seconds.	158
Figure 35 Pulseclarity values for 10 DJ subsets as a function of time in seconds.	159
Figure 36 Eventdensity values for the reference corpus as a function of time in seconds.	160
Figure 37 Eventdensity values for the DJ test corpus as a function of time in seconds.....	160
Figure 38 Chromagrams for the reference corpus.	162
Figure 39 Histogram of the Reference corpus' Mode values.	163
Figure 40 Histogram of HCDF-values for the reference corpus.	164
Figure 41 HCDFover1 and Inharmonicity for the reference corpus.....	165
Figure 42 Loadings plot of PC1 and PC2 for 480 audio files containing more than 30 seconds of sound. ...	167
Figure 43 Correlation values between features.	167
Figure 44 Inharmonicity and Roughness values for the DJs and reference corpus,	169
Figure 45 Zerocross and Brightness values for both the DJs and reference corpus.....	170

Figure 46 Mean Roughness and mean Flux values for both the DJs and reference corpus.....	170
Figure 47 Flux x 10 (red) and Roughness (blue) of the songs in the reference corpus.....	171
Figure 48 Average PC1 and PC2 of the audio files from each act in the dataset.	172
Figure 49 Average statistics for four stages.....	173
Figure 50 Average features for the music played at four stages compared to the reference corpus.	174
Figure 52 Four archetypical shape of DJ sets according to Broughton, & Brewster (2003, 135)	176
Figure 53 A plot of the average “routes” on three different stages.....	177
Figure 54 Average features arranged by sections and stage.....	178
Figure 55 Mean Roughness and mean Pulseclarity for 19 DJ sets from Main Stage	179
Figure 57 Statistics for each cluster.	180
Figure 58. 19 sets from the Main stage mapped according to section and cluster.	180
Figure 59 Average Cluster Length and PC-distance of each stage.	182
Figure 60 The development of some features and throughout Carl Cox’ set, March 27.	184
Figure 61 Average PC-distances and Cluster Length for all DJ sets.	184
Figure 62 Histogram of Difference Zerocross values for 4 different stages.....	186
Figure 63 Percentiles of Difference Zerocross values for 4 different stages.	187
Figure 64 Median Difference values for the points in time where PC distance is above 75.....	188
Figure 65 Roughness over time in Nic Fanciulli.....	190
Figure 66 Prototype of a DJ set on Main Stage.	191
Figure 67 Loadings for Figure 51	205
Figure 68 Loadings for PC 3-4 for DJs and the Reference Corpus	206
Figure 69 Scores for PC 3-4 for DJs and the Reference Corpus	206
Figure 70 Average PC3 and PC4 for sections, and stages.....	207

Bibliography

- Aiden, Erez, and Jean-Baptiste Michel. 2013. *Uncharted, Big Data as a Lens on Human Culture*. New York, New York: Riverhead Books.
- Akpan, Niskan. 2015. "Computer Scientists Prove 80s Pop Music Is Boring." *PBS NewsHour*. <http://www.pbs.org/newshour/updates/computer-scientists-prove-80s-music-boring/>.
- Anagnostopoulou, Christina, and Chantal Buteau. 2010. "Can Computational Music Analysis Be Both Musical and Computational?" *Journal of Mathematics and Music* 4 (2): 75–83. doi:10.1080/17459737.2010.520455.
- Andersen, Jesper Steen. 2014. "Using The Echo Nest's Automatically Extracted Music Features For A Musicological Purpose." In *4th International Workshop on Cognitive Information Processing*. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6844510>.
- Aucouturier, Jean-Julien, and Emmanuel Bigand. 2012. "Mel Cepstrum & Ann Ova: The Difficult Dialog Between MIR and Music Cognition." *ISMIR'12*, 397–402.
- — —. 2013. "Seven Problems That Keep MIR from Attracting the Interest of Cognition and Neuroscience." *Journal of Intelligent Information Systems* 41 (3): 483–97. doi:10.1007/s10844-013-0251-x.
- Balen, J. M. H. van, J. A. Burgoyne, F. Wiering, and R. C. Velkamp. 2013. "An Analysis of Chorus Features in Popular Song." *Proceedings of the 14th Society of Music Information Retrieval Conference (ISMIR)*. <http://dspace.library.uu.nl/handle/1874/281392>.
- Balkwill, Laura-Lee, and William Forde Thompson. 1999. "A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues." *Music Perception: An Interdisciplinary Journal* 17 (1): 43–64. doi:10.2307/40285811.
- Beard, David, and Kenneth Gloag. 2016. *Musicology: The Key Concepts*. 2nd ed. New York, NY, USA: Routledge.
- Bent, Ian, and Anthony Pople. 2017. "Analysis." *Grove Music Online. Oxford Music Online*. Accessed June 13. <http://www.oxfordmusiconline.com.ep.fjernadgang.kb.dk/subscriber/article/grove/music/41862>.
- Bent, Ian, and William Drabkin. 1987. *Analysis*. New York: Norton.
- Berry, David M. 2012. "Introduction." In *Understanding Digital Humanities*, edited by David M. Berry, 1–20. Basingstoke: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230371934.0005>.
- Bertin-Mahieux, Thierry, Daniel P. W. Ellis, Brian Whitman, and Paul Lamere. 2011. "The Million Song Dataset." *ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference*, 591–96.

- Björnberg, Alf. 1996. "Om tonal analys av nutida populärmusik" *Dansk årbog for musikforskning* XXIV: 69–84.
- Bollier, David. 2010. *The Promise and Peril of Big Data*. Washington, DC: Aspen Institute, Communications and Society Program.
- Bourriaud, Nicolas. 2005. *Postproduction: Culture as Screenplay: How Art Reprograms the World*. New York: Lukas & Sternberg.
- Box, George, and Norman Richard Draper. 1987. *Empirical Model-Building and Response Surfaces*. New York: Wiley.
- boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5): 662–79. doi:10.1080/1369118X.2012.678878.
- Brewster, Bill, and Frank Broughton. 2000. *Last Night a Dj Saved My Life: The History of the Disc Jockey*. New York: Grove Press.
- Brewster, Bill, and Frank Broughton. 2010. *The Record Players: DJ Revolutionaries*. Great Britain: djhistory.com.
- Bro, Rasmus, and Age K. Smilde. 2014. "Principal Component Analysis." *Analytical Methods* 6 (9): 2812. doi:10.1039/c3ay41907j.
- Bronson, Bertrand H. 1949. "Mechanical Help in the Study of Folk Song." *The Journal of American Folklore* 62 (244): 81–86. doi:10.2307/536303.
- Brophy, Ben, and Jerry Frempong. 2010. *Everything You Need To Know About DJ'ing & Success: Danny Rampling Shares His 20 Years Experience at the Top*. London: Aurum Press Ltd.
- Broughton, Frank, and Bill Brewster. 2003. *How to DJ Right: The Art and Science of Playing Records*. New York: Grove Press.
- Burgoyne, John. "Stochastic Processes and Database-Driven Musicology." McGill University, 2012. http://digitool.library.mcgill.ca/R/-?func=dbin-jump-full&object_id=107704&silos_library=GEN01.
- Burgoyne, John Ashley, Ichiro Fujinaga, and J. Stephen Downie. 2015. "Music Information Retrieval." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 213–28. Chichester, UK: John Wiley & Sons, Ltd. <http://doi.wiley.com/10.1002/9781118680605.ch15>.
- Burgoyne, John Ashley, Jonathan Wild, and Ichiro Fujinaga. 2011. "An Expert Ground-Truth Set for Audio Chord Recognition and Music Analysis." In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 633–38, http://ismir.net/proceedings/index.php?table_name=papers&function=details&where_field=Id&where_value=1169.
- Butler, Mark Jonathan. 2006. *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music*. Indiana University Press.

- Campbell, Hank. 2012. "The Million Song Dataset - Science Concludes Modern Music Too Loud, All Sounds The Same." *Science* 2.0, July 30.
http://www.science20.com/science_20/million_song_dataset_science_concludes_modern_music_too_loud_all_sounds_same-92614.
- Cannam, C., C. Landone, and M. Sandler. 2010. "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files." In *Proceedings of the ACM Multimedia 2010 International Conference*, 1467–1468.
- Celma, Òscar. 2010. *Music Recommendation and Discovery: The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer Science & Business Media.
- Choudhury, Monojit, Ranjita Bhagwan, and Kalika Bali. 2013. "The Use of Melodic Scales in Bollywood Music: An Empirical Study." In *Proceedings of the 14th International Society for Music Information Retrieval Conference*. International Society for Music Information Retrieval.
<http://research.microsoft.com/apps/pubs/default.aspx?id=198398>.
- CIRMMTvideo. 2012. *J. Stephen Downie - Large-Scale Music Audio Analysis: E-Science or E-Musicology?*
<https://www.youtube.com/watch?v=gfXM1Q7d6Bk>.
- Clarke, Eric, and Nicholas Cook. 2004. *Empirical Musicology*. Oxford University Press.
<http://www.oxfordscholarship.com.ep.fjernadgang.kb.dk/view/10.1093/acprof:oso/9780195167498.01.0001/acprof-9780195167498>.
- Clement, Tanya E. 2012. "Half-Baked: The State of Evaluation in the Digital Humanities." *American Literary History* 24 (4): 876–90.
- Clercq, Trevor de, and David Temperley. 2011. "A Corpus Analysis of Rock Harmony." *Popular Music* 30 (1): 47–70. doi:10.1017/S026114301000067X.
- Cook, Nicholas. 1987. *A Guide to Musical Analysis*. New York: G. Braziller.
- — —. 2005. "Towards the Compleat Musicologist?" *ISMIR, International Conference on Music Information Retrieval 2005*.
- — —. 2010. "The Ghost in the Machine: Towards a Musicology of Recordings." *Musicae Scientiae* 14 (2): 3–21. doi:10.1177/102986491001400201.
- Cook, Nicholas, and Eric Clarke. 2004. "Introduction: What Is Empirical Musicology." In *Empirical Musicology: Aims, Methods, Prospects*, edited by Nicholas Cook and Eric Clarke. New York: Oxford University Press.
- Crawford, Tim, and Lorna Gibson. 2009. *Modern Methods for Musicology: Prospects, Proposals, and Realities*. Digital Research in the Arts and Humanities. Farnham: Ashgate.
- Dalton, Craig, and Jim Thatcher. 2014. "What Does a Critical Data Studies Look Like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data.'" *Society and Space Open Site*. June 2.

- Clercq, Trevor de, and David Temperley. 2011. "A Corpus Analysis of Rock Harmony." *Popular Music* 30, no. 1 : 47–70. doi:10.1017/S026114301000067X.
- Downie, J., D. Byrd, and T. Crawford. 2009. "Ten Years of ISMIR: Reflections on Challenges and Opportunities." In *Proceedings of the 10th Society for Music Information Retrieval Conference (ISMIR)*, 13–18. <http://dc.ofai.at/browser?j=712&y=all>.
- Downie, J. Stephen. 2003a. "Music Information Retrieval." In *Annual Review of Information Science and Technology*, 37, 2003:295–340.
- — —. 2003b. "Toward the Scientific Evaluation of Music Information Retrieval Systems." In *ISMIR 2003 Conference Proceedings*. Johns Hopkins University.
- Downie, J. Stephen, Xiao Hu, Jin Ha Lee, Kahyun Choi, Sally Jo Cunningham, and Yun Hao. 2014. "Ten Years of MIREX: Reflections, Challenges and Opportunities." In *ISMIR 2014*. ISMIR. <http://researchcommons.waikato.ac.nz/handle/10289/8910>.
- Downie, J., Kris West, Andreas Ehmann, and Emmanuel Vincent. 2005. "The 2005 Music Information Retrieval Evaluation Exchange (Mirex 2005): Preliminary Overview." In *6th Int. Conf. on Music Information Retrieval (ISMIR)*, 320–323.
- Dredge, Stuart. 2013. "Pop Music Is Louder, Less Acoustic and More Energetic than in the 1950s." *The Guardian*. <http://www.theguardian.com/technology/2013/nov/25/pop-music-louder-less-acoustic>.
- Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display" *Digital Humanities Quarterly*. 5 (1). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Dutcher, Jenna. 2014. "What Is Big Data?" *Datascience@berkeley*. <http://datascience.berkeley.edu/what-is-big-data/>.
- Echo Nest. 2013a. "People Liked Their Music Fastest in the '80s." *Echo Nest Blog*. <http://blog.echonest.com/post/61686021710/people-liked-their-music-fastest-in-the-80s>.
- — —. 2013b. "The Echo Nest Blog - Tag: Audio Trends over Time." *Echo Nest Blog*. <http://blog.echonest.com/tagged/audio+trends+over+time?og=1>.
- Erickson, Raymond. 1968. "Music Analysis and the Computer." *Journal of Music Theory* 12 (2): 240. doi:10.2307/843312.
- Esparza, Tlacacl Miguel, Juan Pablo Bello, and Eric J. Humphrey. 2015. "From Genre Classification to Rhythm Similarity: Computational and Musicological Insights." *Journal of New Music Research* 44 (1): 39–57. doi:10.1080/09298215.2014.929706.
- Essid, B, G. Richard, and B. David. 2006. "Instrument Recognition in Polyphonic Music Based on Automatic Taxonomies." *IEEE Transactions on Audio, Speech, and Language Processing*. <http://hal-institut-mines-telecom.archives-ouvertes.fr/hal-00477670>.
- Fell, Michael, and Caroline Sporleder. 2014. "Lyrics-Based Analysis and Classification of Music." *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 620–31.

- Fikentscher, Kai. 2000. *"You Better Work!": Underground Dance Music in New York City*. Hanover, NH: University Press of New England.
- — —. 2002. "The Disc Jockey as Composer, Or, How I Became a Composing DJ." *Current Musicology*, no. 67/68: 93–98.
- — —. 2013. "'It's Not the Mix, It's the Selection': Music Programming in Contemporary DJ Culture." In *DJ Culture in the Mix: Power, Technology, and Social Change in Electronic Dance Music*, edited by Anna Gavanis and Hillegonda C. Rietveld, 2013:123.
- Fink, Robert. 2013. "Big (Bad) Data." *Musicology Now*. <http://musicologynow.ams-net.org/2013/08/big-bad-data.html>.
- Finnemann, Niels Ole. 2015. "Humaniora og de Digitale Medier." In *Kampen Om Disciplinerne: Videnskab og Videnskabelighed I Humanistisk Forskning*, edited by D. B. Pedersen, S. Køppe, and F. Stjernfeldt, 1 udgave, 317–41. Copenhagen: Hans Reitzel.
- Foster, Ian. 2011. "How Computation Changes Research." In *Switching Codes*, edited by Thomas Bartscherer, 15. Chicago University Press.
- Gold, Matthew K. 2012. "Day of DH: Defining the Digital Humanities." In *Debates in the Digital Humanities*, 69–71. Minnesota: University of Minnesota Press.
- Greasley, Alinka E., and Helen M. Prior. 2013. "Mixtapes and Turntablism: DJs' Perspectives on Musical Shape." *Empirical Musicology Review* 8 (1): 23–43.
- Gush, Charlotte. 2015. "the Hip Hop Revolution Saved Pop Music in 1991 | Read." *I-D*. May 6. https://i-d.vice.com/en_gb/article/the-hip-hop-revolution-saved-pop-music-in-1991.
- Haas, W. B. de. 2012. "Music Information Retrieval Based on Tonal Harmony." Utrecht University
- Haas, W. Bas de, and Frans Wiering. "Hooked on Music Information Retrieval." *Empirical Musicology Review* 5, no. 4 (2010): 176–85.
- Hadley, Daniel.J. 1993. "'Ride the Rhythm': Two Approaches to DJ Practice." *Journal of Popular Music Studies* 5 (1): 58–67.
- Hallinan, B., and T. Striplas. 2016. "Recommended for You: The Netflix Prize and the Production of Algorithmic Culture." *New Media & Society* 18 (1): 117–37. doi:10.1177/1461444814538646.
- Hauger, David, Markus Schedl, Andrej Košir, and Marko Tkalčič. 2013. "The Million Musical Tweets Dataset: What Can We Learn From Microblogs." In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*.
- Hayles, Katherine. 2012. "How We Think: Transforming Power and Digital Technologies." In *Understanding Digital Humanities*, edited by David Berry, 42–66. Palgrave Macmillan.
- Herman, Bill D. 2006. "Scratching Out Authorship: Representations of the Electronic Music DJ at the Turn of the 21st Century." *Popular Communication* 4 (1): 21–38. doi:10.1207/s15405710pc0401_3.

- Herremans, Dorien, David Martens, and Kenneth Sørensen. 2014a. "Looking into the Minds of Bach, Haydn and Beethoven: Classification and Generation of Composer-Specific Music." Working Paper 2014001. University of Antwerp, Faculty of Applied Economics. <https://ideas.repec.org/p/ant/wpaper/2014001.html>.
- — —. 2014b. "Dance Hit Song Prediction." *Journal of New Music Research* 43 (3): 291–302. doi:10.1080/09298215.2014.881888.
- Hessen, Arjan Van. 2014. "Infrastructures: Will They Be Used?" Slides from Digital Humanities Conference, Lausanne, Switzerland. <https://avindh2014.files.wordpress.com/2014/06/keynote-vanhessen.pdf>.
- Hoffman, Miles. 2005. *The NPR Classical Music Companion : An Essential Guide for Enlightened Listening*. Lightning Source Inc.
- Honingh, Aline, J. Ashley Burgoyne, Peter van Kranenburg, and Anja Volk. 2014. "Strengthening Interdisciplinarity in MIR: Four Examples of Using MIR Tools for Musicology." *ILLC Publications, Prepublication Series* PP 2014-18. <http://dspace.library.uu.nl/handle/1874/303786>.
- Huron, David. 1995. *The Humdrum Toolkit: Reference Manual*. Center for Computer Assisted Research in the Humanities.
- — —. 1996. "The Melodic Arch in Western Folksongs." *Computing in Musicology* 1996 (10): 3–23.
- — —. 1999. "The New Empiricism: Systematic Musicology in a Postmodern Age.", *Berkeley, University of California*.
- — —. 2013. "On the Virtuous and the Vexatious in an Age of Big Data." *Music Perception: An Interdisciplinary Journal* 31 (1): 4–9. doi:10.1525/mp.2013.31.1.4.
- Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.
- Indaco, Agustin, and Lev Manovich. 2016. "Urban Social Media Inequality: Definition, Measurements, and Application," *arXiv preprint arXiv:1607.01845*. <https://arxiv.org/abs/1607.01845>.
- Jehan, T, and D DesRoches. 2014. "The Echo Nest Analyzer Documentation." <https://www.yumpu.com/en/document/view/35577894/analyzer-documentation-the-echo-nest-developer-center>.
- Jockers, Matthew Lee. 2013. *Macroanalysis, Digital Methods and Literary History*. Urbana, IL: University of Illinois Press.
- Kell, Thor, and George Tzanetakis. 2013. "Empirical Analysis of Track Selection and Ordering in Electronic Dance Music Using Audio Feature Extraction." In *ISMIR 2013*, 505–10. http://ismir2013.ismir.net/wp-content/uploads/2013/09/210_Paper.pdf.
- Kirschenbaum, M. G. 2009. Interview with Katherine Hayles. College Park, MD and Hillsborough, NC: 9 January 2009.

- Kranenburg, Peter van, J. Garbers, A. Volk, F. Wiering, L. P. Griep, and L. C. Velkamp. 2007. "Towards Integration of MIR and Folk Song Research." *Proceedings of the 8th International Conference on Music Information Retrieval*. Austrian Computer Society.
- Lamere, Paul. 2009. "Hottt or Nottt?" *Music Machinery*. December 9.
<https://musicmachinery.com/2009/12/09/a-rising-star-or/>.
- Langlois, Tony. 1992. "Can You Feel It?: DJs and House Music Culture in the UK." *Popular Music: A Year Book* 11 (2): 229–38.
- Lartillot, Olivier. 2014. "MIRtoolbox 1.6.1 - User's Manual." Aalborg University, Denmark - Department of Architecture, Design and Media Technology.
<https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox1.6.1guide>.
- Lartillot, Olivier, Olivier Lartillot, Tuomas Eerola, Petri Toiviainen, and Jose Fornari. 2008. "Multi-Feature Modeling of Pulse Clarity: Design, Validation, and Optimization." In *Proceedings of the International Symposium on Music Information Retrieval*.
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.212.5242>.
- Lartillot, Olivier, Petri Toiviainen, and Tuomas Eerola. 2008. "A Matlab Toolbox for Music Information Retrieval." In *Data Analysis, Machine Learning and Applications*, edited by Christine Preisach, Professor Dr Hans Burkhardt, Professor Dr Lars Schmidt-Thieme, and Professor Dr Reinhold Decker, 261–68. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-78246-9_31.
- Laurier, C, J. Grivolla, and P. Herrera. 2008. "Multimodal Music Mood Classification Using Audio and Lyrics." http://mtg.upf.edu/files/publications/Laurier_Grivolla_Herrera_Mood_Lyrics_ICMLA.pdf.
- Lee, Jin Ha, M Cameron Jones, and J. Stephen Downie. 2009. "An Analysis of ISMIR Proceedings: Patterns of Authorship, Topic, and Citation." In *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 57–62. <http://ismir2009.ismir.net/proceedings/PS1-1.pdf>.
- Leman, Marc. 2008. "Systematic Musicology at the Crossroads of Modern Music Research." In *Systematic and Comparative Musicology: Concepts, Methods, Findings*, edited by Albrecht Schneider, 89–115. Frankfurt Am Main: Peter Lang.
- Lerch, Alexander. 2012. *An Introduction to Audio Content Analysis Applications in Signal Processing and Music Informatics*. Hoboken, N.J.: Wiley. <http://onlinelibrary.wiley.com/book/10.1002/9781118393550>.
- Li, Tao, Mitsunori Ogihara, and George Tzanetakis. 2012. *Music Data Mining*. Boca Raton: CRC Press.
- Lieken, Anthony. 2007. "Data Mining Musical Profiles." anthony.lieken.net,
<http://anthony.lieken.net/index.php/Computers/DataMining>.
- Lin, Yu Wei. 2012. "Transdisciplinarity and Digital Humanities." In *Understanding Digital Humanities*, edited by David M. Berry. New York: Palgrave Macmillan.

- Manovich, Lev. 2009. "How to Follow Global Digital Cultures, or Cultural Analytics for Beginners." *Deep Search: The Politics of Search beyond Google*. http://manovich.net/content/04-projects/060-how-to-follow-global-digital-cultures/59_article_2009.pdf.
- — —. 2011. "Trending: The Promises and the Challenges of Big Social Data." <http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data>
- — —. 2012. "How to Compare One Million Images?" In *Understanding Digital Humanities*, edited by David M. Berry. Palgrave Macmillan.
- — —. 2015. "The Science of Culture? Social Computing, Digital Humanities, and Cultural Analytics." <http://manovich.net/index.php/projects/cultural-analytics-social-computing>.
- Marsden, Alan. 2009. "What Was the Question?: Music Analysis and the Computer." In *Modern Methods for Musicology. Prospects, Proposals, and Realities*, edited by Tim Crawford and Lorna Gibson.
- — —. 2012. "Counselling a Better Relationship between Mathematics and Musicology." *Journal of Mathematics & Music* 6 (2): 145–53.
- — —. 2016. "Music Analysis by Computer: Ontology and Epistemology." In *Computational Music Analysis*, edited by David Meredith, 3–28. Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-25931-4_1.
- Mauch, Matthias. 2015. *The Hidden History of Pop Music*. Ted Talk. <http://tedxtalks.ted.com/video/The-Hidden-History-of-Pop-Music>.
- Mauch, Matthias, Robert MacCallum, Mark Levy, and Armand Leroi. 2015. "The Evolution of Popular Music: USA 1960–2010." *Royal Society Open Science* 2 (5): 150081. doi:10.1098/rsos.150081.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- McCarty, Willard. 2007. "Knowing ... : Modeling in Literary Studies." In *Companion to Digital Literary Studies*, edited by Susan Schreibman and Ray Siemens, 391–401. Oxford: Blackwell.
- McKay, Cory. 2010. "Automatic Music Classification with jMIR." PhD thesis, McGill University.
- McVicar, M., R. Santos-Rodriguez, Yizhao Ni, and Tijl De Bie. 2014. "Automatic Chord Estimation from Audio: A Review of the State of the Art." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22 (2): 556–75. doi:10.1109/TASLP.2013.2294580.
- Mendel, Arthur. 1962. "Evidence and Explanation." In *International Musicological Society Report of the Eighth Congress New York I*, edited by Jan LaRue, 3–18. Basel: Bärenreiter.
- — —. 1969. "Some Preliminary Attempts at Computer-Assisted Style Analysis in Music." *Computers and the Humanities* 4 (1): 41–52.
- Meyer, Leonard B. 1989. *Style and Music: Theory, History, and Ideology*. Studies in the Criticism and Theory of Music. Philadelphia: University of Pennsylvania Press.

- Michel, Jean-Baptiste, Martin A. Nowak, Erez Lieberman Aiden, Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–182. doi:10.1126/science.1199644.
- Montano, Ed. 2009. *DJ Culture in the Commercial Sydney Dance Music Scene*. Dancecult. <http://researchbank.rmit.edu.au/view/rmit:19110>.
- Moretti, Franco. 2000. "Conjectures on World Literature." *New Left Review* 1. <http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>.
- — —. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.
- — —. 2011. "Network Theory, Plot Analysis." *New Left Review*, II, , no. 68: 80–102.
- — —. 2013. *Distant Reading*. London: Verso. <http://www.amazon.de/Distant-Reading-Franco-Moretti/dp/1781680841>.
- Müller, M. 2015. *Fundamentals of Music Processing*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-21945-5>.
- Müller, M., D. P. W. Ellis, A. Klapuri, and G. Richard. 2011. "Signal Processing for Music Analysis." *IEEE Journal of Selected Topics in Signal Processing* 5 (6): 1088–1110. doi:10.1109/JSTSP.2011.2112333.
- Nattiez, Jean-Jacques. 1975. *Fondements D'une Semiologie de La Musique*. Paris: Union Generale d'Editons.
- Nettheim, N. 1997. "A Bibliography of Statistical Applications in Musicology." *Musiology Australia* 20: 94–106.
- Newbould, Brian. 2017. "Texture." Edited by Alison Latham. *The Oxford Companion to Music*. *Oxford Music Online*. Accessed April 26. <http://www.oxfordmusiconline.com.ep.fjernadgang.kb.dk/subscriber/article/opr/t114/e6743>.
- Ørmen, J., and A. M. Thorhauge. 2015. "Smartphone Log Data in a Qualitative Perspective." *Mobile Media & Communication* 3 (3): 335–50. doi:10.1177/2050157914565845.
- Plumbley, Mark D., Samer A. Abdallah, Juan Pablo Bello, Mike E. Davies, Giuliano Monti, and Mark B. Sandler. 2002. "Automatic Music Transcription and Audio Source Separation." *Cybernetics & Systems* 33 (6): 603–627.
- Porsdam, Helle. 2013. "Digital Humanities: On Finding the Proper Balance between Qualitative and Quantitative Ways of Doing Research in the Humanities" 7 (3). <http://www.digitalhumanities.org/dhq/vol/7/3/000167/000167.html>.
- Ramsay, Stephen. 2013. "Who's in and Who's out." In *Defining Digital Humanities : A Reader*, edited by Melissa Terras, Julianne Nyhan, and Edward Vanhoutte. Farnham, Surrey, England: Ashgate Publishing Limited ; Burlington, VT : Ashgate Publishing Company.
- Reynolds, Simon. 2012. "How Rave Music Conquered America." *The Guardian*, August 2. <http://www.theguardian.com/music/2012/aug/02/how-rave-music-conquered-america>.

- Rieder, Bernhard, and Theo Röhle. 2012. "Digital Methods." In *Understanding Digital Humanities*, edited by David M. Berry. Vol. 2012. Palgrave Macmillan.
- Rietveld, Hillegonda C. 2013. "Journey to the Light? Immersion, Spectacle and Mediation." In *DJ Culture in the Mix: Power, Technology, and Social Change in Electronic Dance Music*, edited by Hillegonda C. Rietveld, Bernardo Attias, and Anna Gavanas. New York: Bloomsbury Academic.
- Rodriguez Zivic, P. H., F. Shifres, and G. A. Cecchi. 2013. "Perceptual Basis of Evolving Western Musical Styles." *Proceedings of the National Academy of Sciences* 110 (24): 10034–38. doi:10.1073/pnas.1222336110.
- Rogers, Richard. 2013. *Digital Methods*. Cambridge, Massachusetts: The MIT Press.
- Scheirer, E. D. 1998. "Tempo and Beat Analysis of Acoustic Musical Signals." *The Journal of the Acoustical Society of America* 103 (1): 588–601.
- Schellenberg, Glenn E., and Christian von Scheve. 2012. "Emotional Cues in American Popular Music: Five Decades of the Top 40." *Psychology of Aesthetics, Creativity, and the Arts* 6 (3): 196–203. doi:10.1037/a0028024.
- Schnapp, Jeffrey, Peter Lunenfeld, and Todd Presner. 2009. "Digital Humanities Manifesto 2.0." <http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/>.
- Schnitzer, Dominik, Arthur Flexer, and Gerhard Widmer. 2009. "A Filter-and-Refine Indexing Method for Fast Similarity Search in Millions of Music Tracks." In *Proceedings of the 10th International Society for Music Information Retrieval Conference*. Kobe, Japan.
- Schöch, Christof. 2013. "Big? Smart? Clean? Messy? Data in the Humanities." *Journal of Digital Humanities*. November 22. <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- Serrà, Joan, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll Arcos. 2012. "Measuring the Evolution of Contemporary Western Popular Music." *Scientific Reports* 2 (July). doi:10.1038/srep00521.
- Sethares, William A. 1998. *Tuning, Timbre, Spectrum, Scale*. London ; New York: Springer.
- Shanin, Teodor. 1972. "Models in Thought." In *The Rules of the Game: Cross-Disciplinary Essays on Models in Scholarly Thought*, edited by Teodor Shanin, 1–22. London: Tavistock.
- Sherburne, Philip. 2012. "Top 100 DJs Poll: Who Won, Who Lost, and What the Hell Is Hardstyle? | SPIN." October 23. <http://www.spin.com/2012/10/top-100-djs-poll-who-won-who-lost-and-what-the-hell-is-hardstyle/>.
- Small, Christopher. 1998. *Musicking: The Meanings of Performing and Listening*. Middletown, Conn.: Wesleyan University Press.
- Smith, John R. 2013. "Riding the Multimedia Big Data Wave." In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1–2. SIGIR '13. New York, NY, USA: ACM. doi:10.1145/2484028.2494492.

- Sturm, Bob L. 2012. "Paper of the Day (Po'D): Semantic Gap?? Schemantic Schmap!! Edition" *Pursuits in the Null Space*. http://media.aau.dk/null_space_pursuits/2012/05/paper-of-the-day-pod-semantic-gap-schemantic-schmap-edition.html.
- — —. 2013. "Reviews: Publications: Alexander Lerch: 'An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics.'" *Computer Music Journal* Vol. 37 (4): pp.90-91.
- Sturm, Bob L., and Nick Collins. 2014. "The Kiki-Bouba Challenge: Algorithmic Composition for Content-Based MIR Research & Development." In *Proceedings of ISMIR 2014, Taipei, Taiwan*.
- Sundram, Jason. 2010. "Danceability and Energy: Introducing Echo Nest Attributes." *Running with Data - Fun with Data, Music, and Visualization*. <http://runningwithdata.com/post/1321504427/danceability-and-energy>.
- Swartz, A. 2002. "MusicBrainz: A Semantic Web Service." *IEEE Intelligent Systems* 17 (1): 76–77. doi:10.1109/5254.988466.
- Temperley, David, and Trevor de Clercq. 2013. "Statistical Analysis of Harmony and Melody in Rock Music." *Journal of New Music Research* 43 (2): 187-204. doi:10.1080/09298215.2013.839525.
- The Echo Nest blog. 2011. "Speechiness - Is It Banjo or Banter." *The Echo Nest Blog*. <http://blog.echonest.com/post/12324116771/speechiness-is-it-banjo-or-banter>.
- Thompson, Derek. 2015. "1991: The Most Important Year in Pop-Music History." *The Atlantic*, May 8. <https://www.theatlantic.com/entertainment/archive/2015/05/1991-the-most-important-year-in-music/392642/>.
- Tukey, John W. 1962. "The Future of Data Analysis." *The Annals of Mathematical Statistics* 33 (1): 1–67. doi:10.1214/aoms/1177704711.
- Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, Mass.; Menlo Park, Calif.; London; Amsterdam: Addison-Wesley.
- Tzanetakis, G., and P. Cook. 2002. "Musical Genre Classification of Audio Signals." *IEEE Transactions on Speech and Audio Processing* 10 (5): 293–302. doi:10.1109/TSA.2002.800560.
- Tzanetakis, George. 2014. *UVic MIR Course - Audio Feature Extraction*. <https://www.youtube.com/watch?v=cd2Jfi0PE2Y&list=PLqjF740IZaFSGRWENDrsC2bQ3rhZlWJLM&index=1>.
- Vigen, Tyler. 2015. *Spurious Correlations*. New York: Hachette Books.
- Viro, Vladimir. 2011. "Peachnote: Music Score Search and Analysis Platform" International Society for Music Information Retrieval Conference (ISMIR).
- Volk, A., F. Wiering, and P. Kranenburg van. 2011. "Unfolding the Potential of Computational Musicology." In *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations*, 37–44. Leeuwarden: Fryske Akademy. http://www.staff.science.uu.nl/~fleis102/CompMus_Volketal.pdf

- Volk, Anja, and W.B. de Haas. 2013. "A corpus-based study on ragtime syncopation." In *Proceedings of the International Society for Music Information Retrieval Conference*. <http://hdl.handle.net/1874/289635>.
- Volk, Anja, and Aline Honingh. 2012. "Mathematical and Computational Approaches to Music: Challenges in an Interdisciplinary Enterprise." *Journal of Mathematics and Music* 6 (2): 73–81. doi:10.1080/17459737.2012.704154.
- Vos, Piet G., and Jim M. Troost. 1989. "Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance." *Music Perception: An Interdisciplinary Journal* 6 (4): 383–96. doi:10.2307/40285439.
- Wakefield, Gregory H. 1999. "Mathematical Representation of Joint Time-Chroma Distributions" In *Proc. SPIE 3807, Advanced Signal Processing Algorithms, Architectures, and Implementations IX*, 637, edited by Franklin T. Luk, 637–45. doi:10.1117/12.367679.
- Wallmark, Zachary. 2013. "Big Data and Musicology: New Methods, New Questions." *Proceedings of the National Academy of Sciences* 10 (1073).
- Walser, Robert. 2003. "Popular Music Analysis: Ten Apothegms and Four Instances." In *Analyzing Popular Music*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511482014.002>.
- Whitman, Brian. 2013. "How Music Recommendation Works – and Doesn't Work." *Brian Whitman*. <http://notes.variogr.am/post/37675885491/how-music-recommendation-works-and-doesnt-work>.
- Wiering, Frans. 2007. "Can Humans Benefit from Music Information Retrieval?" In *Proceedings of the 4th International Conference on Adaptive Multimedia Retrieval: User, Context, and Feedback*, 82–94. AMR'06. Berlin, Heidelberg: Springer-Verlag. <http://dl.acm.org/citation.cfm?id=1756944.1756954>.
- — —. 2009. "Meaningful Music Retrieval." In *1st Workshop on the Future of Music—Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*.
- — —. 2012. "Balancing Computational Means and Humanities Ends in Computational Musicology." <http://www.staff.science.uu.nl/~wieri103/presentations/DHlectureWieringDecember2012.pdf>.
- Wiering, Frans, and Emmanouil Benetos. 2013. "Digital Musicology and MIR: Papers, Projects and Challenges." <http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2014/02/lbd4.pdf>
- Wilkins, Matthew. 2012. "Canons, Close Reading, and the Evolution of Method." In *Debates in the Digital Humanities*. <http://dhdebates.gc.cuny.edu/debates/part/5>

Acknowledgements

(I will switch to Danish here, as it feels very odd and impersonal to say thanks in a foreign language to persons who know Danish)

Først og fremmest tak til Danmark, eller borgerne i det her land. Det var jer der betalte min stilling, det her projekt. Jeg håber, at det kommer til at vise sig at blive *value for money*. I den allerbredeste forståelse af ordet værdi.

Tak til min hovedvejleder Jack Andersen. For konstruktiv og empatisk vejledning, som stikker meget dybere end de her tre år. Du har guidet mig til min vigtigste arbejdsmæssige erkendelse de sidste 10 år: At jeg giver mest til en arbejdsplads, hvis jeg giver noget af mig selv, frem for hvis jeg giver det, som jeg tror, andre vil have. Det var præcis det, jeg havde brug for at vide.

Tak til bivejleder Morten Michelsen. Også for konstruktiv vejledning og for at få mig til stå ved.

Tak til Henrik Smith-Sivertsen. For titlen på afhandlingen og kapitel 6. Og for omsorg og snakke om alt muligt forskningsrelateret.

Kæmpe tack til Sune Bechmann og Anamaria Dutceac. Uundværlig sparring og hjælp, da jeg i sin tid formulerede projektet.

Tak til CoSound. Især Bjørn og Jens. Fordi I tog godt imod mig på DTU i mit ophold. Og tak for alt hjælpsomhed, og fordi I forklarede mig alt basic om MIR og gav mig et godt indblik i ingeniørfeltet. Og tak for at I skubbede mig ud i at bruge MATLAB, selv om det betød, at jeg hele tiden skulle spørge jer om alt muligt.

Tak til Trine Schreiber. For din tilgængelighed og imødekommenhed, og rådgivning om det at være ph.d.

Tak til ph.d. kollegaer på IVA. Til de gamle (Rikke, Ole, Sarah, Lorna, Niels Peder og Sille) og til generationen efter mig (Maria, Lisa, Martin, Eva, Susanne, Mia, Yu-Tsu). Livet som ph.d. er så meget federe med nogle at dele det med. Og især nogen som gider at

sidde snakke zombie-/mumieteori, eller forestille sig en tv-dokumentar om gravide, der ikke vidste de er gravide. (Ja, man skulle nok have været der).

Tak til Jacob Thøgersen, Bente Larsen, Torben Sangild og Thorkild Hanghøj. For sparring og moralsk opbakning.

Og tak til min familie. Til Silje, Norbert og Dikte, min mor og min far. For det hele. I har alle min dybeste respekt.

Summary

This thesis investigates how musicologists can use digital audio content analysis (ACA) methods for analyzing large amounts of music. Abundances of music have been digitized in recent years, and the field of Music Information Retrieval has created ACA methods to automatically calculate the so-called features from audio files, such as tempo or tonality. However, the vast majority of musicologists do not yet use ACA tools for analyses, although they can be a means to deal with these abundances of audio files.

I illuminate my research question by seeking inspiration in the broader field of the Digital Humanities for discussing how to apply quantitative techniques within a qualitative field. I examine ACA methods' practical and epistemological value for musicologists in three case studies: In case 1, I examine machine-learned metrics that allows the creation of features that estimate intuitive qualities of the music. In Case 2, I discuss the epistemological value of an existing big data music analysis. And in Case 3, I apply ACA methods for assisting an analysis of 89 DJ sets played at a recent festival.

Data can be a means to “enhance the perception”. I.e. to observe things that otherwise would have been hard to find. Musicologists can apply data techniques for posing new questions, for example about many pieces of music, or about musical aspects adhering to a subset of a corpus. They can apply them for a “second” empirical opinion on the corpus. With modern software, digital methods allow swift exploration of musical aspects in large corpora for example through visualization, such as to create crude maps of the music. Statistics enable us to generalize, find larger patterns, but also to nuance and find individual differences.

Music analytically, ACA features can both represent new and traditional ways of measuring music. However, there seems to be a current ceiling of about 70-85% correctness for ACA tasks, and this affects the conclusions for large-scale analysis. There is also a current tendency that timbral aspects play a more prominent role than previously. However, the features are created by opaque and complicated algorithms. This impedes the translation from feature to music analysis, and the analysis becomes imprecise or uncertain on a qualitative level, despite exact methods. If musicologists want to exploit the potentials of ACA methods, they will have to get used to not understanding the connection between features and music entirely.

Resumé

I denne afhandling undersøger jeg, hvordan musikvidenskaben kan bruge digitale *audio content analysis* (ACA)-metoder til at analysere større mængder musik, end man ville gøre manuelt. I de seneste år er store mængder musik blevet digitaliseret, og samtidig er der udviklet en række ACA-metoder, der vha. algoritmer kan udregne såkaldte features, som estimerer aspekter af musik, såsom tempo, toneart, osv. Men musikvidenskaben anvender stort set ikke ACA-metoder.

Jeg belyser mit spørgsmål ved dels at teoretisk med afsæt i især digital humaniora at diskutere, hvordan kvantitative metoder kan anvendes i en ellers typisk kvalitativ disciplin. Og dels undersøger jeg ACA-metoders praktiske og epistemologiske værdi ud fra 3 cases: I case 1 undersøger jeg maskinlærte features, der estimerer musiks intuitive kvaliteter. I case 2 undersøger jeg en allerede eksisterende musikalsk big data-analyse. Og I case 3 bruger jeg ACA-metoder til at assistere en analyse af 89 DJ sets.

Jeg konkluderer, at data kan være middel for musikforskere til at udvide perceptionen. Dvs. til at få øje på ting, der ville være svære at få øje på ellers. Musikforskere kan bruge teknikkerne til at stille nye spørgsmål, som kan undersøges med hjælp fra data. Det kunne fx være spørgsmål om mange stykker musik, eller spørgsmål af typen *hvilke musikalske karakteristika er der i delmængde A, men ikke i B?* Svarene på disse spørgsmål kan indgå som led i kvalitative spørgsmål, for moderne software muliggør hurtige datavisualiseringer, som gør at vi kan udforske musikalske aspekter i store korpora eller kreere kort over musikken. Statistik kan hjælpe til at generalisere, finde mønstre, men også med at nuancere og finde individuelle forskelle.

ACA features kan både repræsentere nye musikalske mål, men samtidig også simulere traditionelle musikanalytiske mål. Men der synes at være et loft på omkring 70-85% korrekthed, hvad angår de traditionelle metoder. Det påvirker, de konklusioner der kan drages om mange stykker musik. Der er for øjeblikket også en generel tendens til at især klanglige aspekter, såsom masterering, spiller en større rolle end i tidligere musikanalyse. Features er samtidig også skabt ud fra komplicerede og uigennemsigtige algoritmer, og det besværliggør oversættelsen fra feature til musikanalyse. Dermed bliver analysen kvalitativt upræcis eller usikker. På trods af ellers præcise metoder. Hvis musikforskere vil udnytte fordelene ved ACA-metoder – og det er der mange gode grunde til – bliver de også nødt til at indstille sig på aldrig helt at kunne forstå sammenhængen mellem musik og feature.